PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/12, 15/13, C07K 13/00
A61K 37/02

A1

(11) International Publication Number: WO 92/01787

(43) International Publication Date: 6 February 1992 (06.02.92)

(21) International Application Number: PCT/GB91/01253

(22) International Filing Date: 25 July 1991 (25.07.91)

(30) Priority data:
9016299.1 25 July 1990 (25.07.90) GB

(71) Applicant *(for all designated States except US)*: CAMBRIDGE ANTIBODY TECHNOLOGY LIMITED [GB/GB]; The Daly Research Laboratories, Babraham Hall, Cambridge, Cambridgeshire CB2 4AT (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : JOHNSON, Kevin, Stuart [GB/GB]; 32 Holmehill, Godmanchester, Cambridgeshire PE18 8EX (GB). JACKSON, Ronald, Henry [GB/GB]; 31 Kingston Street, Cambridge CB1 2NU (GB). CHISWELL, David, John [GB/GB]; 1 Sandhill House, Middle Claydon, Buckingham MK18 2LD (GB).

(74) Agent: STUART, I., A.; Mewburn Ellis, 2 Cursitor Street, London EC4A 1BQ (GB).

(81) Designated States: AT, AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH, CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE, DE (European patent), DK, DK (European patent), ES, ES (European patent), FI, FR (European patent), GA (OAPI patent), GB, GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU, LU (European patent), MC, MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL, NL (European patent), NO, PL, RO, SD, SE, SE (European patent), SN + (OAPI patent), SU, TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.

(54) Title: BINDING DOMAINS

(57) Abstract

This invention concerns binding domains e.g. single chain variable domains which are synthetic analogues of other single chain variable domains of members of an immunoglobulin family or superfamily. In the analogue, one or more amino acid residues is altered as compared to the other domain, so that the analogue is more hydrophilic than the natural domain. The altered amino acid is substituted with a residue which occurs in an analogous position in a member of an immunoglobulin family or superfamily. This increased hydrophilicity means that the synthetic analogue will show less non-specific binding than the natural domain. The analogue may retain the binding specificity of the natural domain. Alternatively, the complementarity determining regions may be altered to change the binding specificity. The invention also concerns methods for making these binding domains.

+ See back of page

+ DESIGNATIONS OF "SU"

It is not yet known for which States of the former Soviet Union any designation of the Soviet Union has effect.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU+	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		

BINDING DOMAINS

The present invention relates to molecules comprising binding domains and methods for their preparation and use. In particular the present invention relates to molecules comprising domains which are synthetic analogues of a natural single variable domain of a member of an immunoglobulin family or superfamily. The present invention also relates to methods for designing said molecules comprising domains, the molecules so designed, and kits and methods for use of said molecules in therapy and diagnosis.

Antibodies and other members of the immunoglobulin superfamily, such as T cell receptors, have the ability to recognise molecules eg. antigens, specifically and bind them with high affinity. In naturally occurring antibodies, the binding site for antigen is formed by the juxtaposition of variable (V) domains from both heavy (H) and light (L) chains. Within each of these chains are three stretches of amino acids, the complementarity determining regions (CDRs), which comprise the residues which interact with antigens. The three CDRs alternate with four framework regions (FR). Winter et al. demonstrated that a single V domain was able to bind antigen with high affinity and specificity [Ward et al. Nature 341, 544-546 (1989)]. It was proposed that these single domain antibodies (VH) would have advantages for

several applications owing to their small size (relative molecular mass (rmm)=13,000) compared to whole antibodies (rmm=160,000) and other antibody fragments.

However, VH domains have unique disadvantages that disappointingly limit their utility. The difficulties encountered reflect at least two properties of VH domains which may be related. They are expressed in low quantities when cloned in bacteria (ca. 200μg/ml of culture supernatant compared to 10mg/ml for Fv fragments) and during purification of VH domains, substantial amounts of material are lost. Concentration of the VH single domain, for example using ultrafiltration and purification on chromatography columns, often leads to poor recoveries. This probably reflects non-specific binding to surfaces. This has been observed directly by Ward et al. (1989) supra., who isolated a high proportion of single-domain antibodies exhibiting significant (and non-specific) binding to plastic. *In vivo*, non-specific binding in tissues would lead to poor performance in applications such as tumour imaging or cancer therapy studies.

These difficulties are not generally experienced with whole antibodies or with fragments of antibodies such as Fv or Fab fragments. Therefore, the problems appear to be a characteristic of antibody fragments containing unpaired single domains.

Thus, the present invention seeks to ameliorate any of the above or other problems associated with single

variable domain binding members, whilst retaining antigen binding for their full commercial exploitation.

The present applicant has realised that the most likely cause of the unfavourable properties of single domain antibodies is the exposure to aqueous solvent, of the hydrophobic face of a single variable domain eg. the VH single domain. In native antibodies, this face interacts with the adjacent hydrophobic face of the VL domain and is buried within the antibody molecule. Exposure of the face would lead to strong interactions with surfaces, for example chromatography matrices, from which material could not be recovered. Additionally, the present applicant believes that exposure of the face to aqueous solvent may lead to decreased stability of the single variable domain eg. the VH single domain, leading to turn to unfolding and loss of binding activity during purification procedures. Most importantly, this hydrophobic face is a potential source of non-specific binding, which considerably limits the utility of these single variable domain molecules in vivo and in vitro.

Although the VH single domain is quite small, its activity depends on a complex tertiary structure comprising interactions between disparate parts of the molecule [Chothia, C. et al. J. Mol. Biol. 186, 651-663 (1985)]. In several cases, there are side-chains in the framework which have important interactions with those in the antigen combining site. For example, in the anti-progesterone antibody DB3, tryptophan 47 (framework) has

been shown to contact progesterone (Arevalo, J.H., Taussig, M. and Wilson, I. personal communication), and -hydroxy vitamin K in the -hydroxy vitamin K-binding antibody IgG1 NEW [Twining, S.S. and Atassi, M.Z. J. Biol. Chem. 253, 5259 (1978)]. In some antibodies, interactions of residue 71 (framework) with CDR residues have been shown to be important for the maintenance of antigen binding [Chothia, C. et al. Nature 342, 877-883 (1989)]. It will be apparent to those skilled in the art, that any modifications to the basic antibody or variable domain structure will most likely have pleiotropic consequences on the tertiary structure of the antibody or variable domain and render them unable to bind antigen with the desired specificity and affinity. This is exemplified in recent experiments with anti-lysozyme single-domain antibodies. A single amino acid change, substituting Asn 35 with His, was found to improve the expression level in E.coli by about 1000 fold. Critically however, the altered molecule was found to bind lysozyme only very weakly (E.S.Ward, L.Reichmann, G.P.Winter, personal communication). Because the interactions responsible for assembly of the active structure, particularly of VH domains, are not completely understood, any amino acid changes made to the basic structure must be selected extremely carefully. Selection of any changes is further complicated by the fact that the residues at the interface of the VH single domain with the VL domain, come from the CDRs as well as

the framework. The framework residues (for example, residues 37,39,45,47,91,93 and 103) at the interface of the anti-lysozyme antibody D 1.3 [Amit, A.G. et al, Science 233 747-753 (1986), Chothia, C. et al., Science 233, 755-758 (1986)] are highly conserved in antibodies of all species. They are invariant or, rarely, have substitutions which are usually conservative, i.e. replacement of the amino acid with one having similar chemical properties. For example, aliphatic hydrophobic residues are usually substituted by similar aliphatic residues. Thus, if substitutions are made in these framework residues, there is a strong risk of disrupting the structure of the molecule and hence binding of antigen by the antibody or single variable domain eg. VH. The CDR residues vary between antibodies and determine the specificity and affinity of binding to antigen. Since the applicant wishes to retain the capacity to bind a variety of antigens with high affinity and specificity, the ability to vary these CDR residues needs to be retained.

The present invention therefore provides a molecule comprising a binding domain having a polypeptide sequence which is an analogue of part or all of a naturally occurring molecule comprising a binding domain having specificity for a particular binding member, in which analogue one or more amino acids are altered to reduce the hydrophobicity of said analogue as compared to the natural binding substance.

The analogue may have substantially the same binding characteristics as the natural binding substance eg. with respect to specificity, affinity, or avidity. In some cases these characteristics may be improved. The alteration may be any amino acid alteration which reduces hydrophobicity of the polypeptide sequence eg. amino acid substitution, deletion or addition.

The molecule comprising a binding domain may comprise an antibody or other receptor molecule and fragments and derivatives of antibodies and receptor molecules. In particular, the molecule may comprise a single variable domain of the type that is present in an antibody molecule. The alteration may be in a complementarity determining region and/or in a framework region. Preferably, the alteration is in a framework region. Where the alteration reducing hydrophobicity, is in a framework region, the complementarity determining regions may also be altered by way of amino acid substitution, deletion, addition or inversion to alter the specificity and or binding characteristics of the binding substance.

The molecule comprising a binding domain may comprise any one or more of the alterations described in examples 2 to 11.

Thus the present invention provides a single chain variable domain which is a synthetic analogue of another single chain variable domain of a member of an immunoglobulin family or superfamily, and in which

analogue, one or more interface amino acids residues of the domain is altered as compared to the said another domain, in which a said altered amino acid is substituted with a residue which occurs in an analogous position in a member of an immunoglobulin family or superfamily, so that the analogue is more hydrophilic than the said another domain.

The altered amino acid residue may be in a framework region. The altered amino acid residue may be in a complementarity determining region. The synthetic analogue may have essentially the same binding activity as the said another domain. The sequence of a complementarity determining region may be additionally altered by way of amino acid substitution, deletion, addition, or inversion, to alter the specificity and/or binding characteristics of the analogue as compared to the said another domain.

The single chain variable domain may be a synthetic analogue of a single variable immunoglobulin heavy chain domain. In which case, the one or more of the amino acid residues 37,39,45,47,91,93 and 103 may be altered. The amino acid alterations may comprise one or more of the following:

i) substitution of valine 37 with glutamine or threonine;

ii) substitution of glutamine 39 with glutamate;

iii) substitution of leucine 45 with glutamine;

iv) substitution of tryptophan 47 with aspartate or

glycine;

v) substitution of tyrosine 91 with threonine, serine or methionine;

vi) substitution of alanine 93 with serine or glutamate;

vii) substitution of tryptophan 103 with glutamate or tyrosine or threonine;

viii) substitution of valine 37, leucine 45, tryptophan 47, alanine 93 and/or tryptophan 103 with any of asparagine, threonine or serine;

ix) substitution of valine 37 with threonine and glutamine 39 with glutamate and tryptophan 47 with glycine;

x) substitution of tyrosine 91 with serine or methionine and alanine 93 with glutamate and tryptophan 103 with threonine.

The single chain variable domains according to the present invention may be coupled to a further molecular moiety. The further molecular moiety may be an enzymic-, fluorescent-, or radio-label, or a portion of an immunoglobulin.

The present invention also provides diagnostic kits which comprise a single chain variable domain as described above together with one or more ancillary reagents for carrying out the diagnostic test.

The present invention also provides therapeutic composition which comprises at least a single chain variable domain as described above. The composition may

also comprise one or more excipients.

In an aspect of the invention there is provided single chain variable domains eg. VH single chain domain frameworks, with improved properties as indicated above and which allow the generation of specificities for various binding partners by substitution of different CDRs into the new framework. The applicant also provides that the framework residues of molecules comprising single domains previously isolated are substituted to make them more polar. The modified molecule should retain the ability to bind the desired antigen. Preferably, the changes would not render the interface immunogenic when administered to humans.

The present invention also provides a method for making a single chain variable domain which is a hydrophilic synthetic analogue of another single chain variable domain of a member of an immunoglobulin family or superfamily, which comprises:

(i) inspecting the interface regions of a said single chain variable domain to identify hydrophobic amino acid residues; and

(ii) producing a said analgue of said single chain variable domain in (i) in which one or more of said hydrophobic residues is substituted with a less hydrophobic residue which occurs in an analogous position in a member of an immunoglobulin family or superfamily.

The method may comprise:

(a) obtaining the nucleotide sequence encoding one

or more of the identified hydrophobic amino acid residues;

(b) using site directed mutagenesis to alter the nucleotide sequence to introduce a triplet coding for the substitute amino acid,

(c) using the altered nucleotide sequence in a recombinant expression system to express the synthetic analogue.

In the method, more than one amino acid residue may be substituted. The substitute amino acids may be derived from naturally monomeric members of the immunoglobulin superfamily. The naturally monomeric member may be Thy-1. The synthetic analogue may have essentially the same binding activity as the said another domain.

Where the molecule comprising a binding domain is an immunoglobulin, or a fragment or derivative of an immunoglobulin, an amino acid site suitable for alteration may be identified by:

i) examining the molecule for hydrophobic amino acids expected to be on the surface of the binding substance,

ii) more specifically examining the amino acid residues related to those expected to be buried at the interface of the heavy and light chain domains of an immunoglobulin,

iii) examining those residues identified in (ii) that would be exposed to the solvent when the single domain is used separately.

An appropriate alteration to make may be identified with reference to the homologous amino acid sequences of members of a family of related substances. For example, the amino acid sequence of a molecule at the site for alteration identified as described above, may be altered to make it homologous at that site in one or more members of the family group.

The family of related substances may comprise a family of immunoglobulins, fragments and derivatives thereof. Alternatively, the family of related substances may comprise a family of proteins containing domains structurally related to the immunoglobulins ie. an immunoglobulin superfamily.

The nucleotide sequence may be altered by site directed mutagenesis using oligonucleotides designed to introduce the required alterations. Alternatively, the alteration may be achieved by use of the technique known as polymerase chain reaction.

The present invention also embraces kits having molecules comprising binding domains as herein provided. The kits may be diagnostic, purification or catalysis kits. The present invention further embraces pharmaceutical products which contain molecules comprising binding domains according to the present invention.

TERMINOLOGY

Binding Domain

This describes an area on the surface of a protein,

or a cavity which specifically binds to, and is therefore defined as complementary with a particular spatial and polar organisation of another molecule. A domain is folded within itself and independently of other parts of the same protein and independently of a complementary binding member.

Immunoglobulin

This describes a group of structurally related proteins consisting of two pairs of heavy polypeptide chains and two pairs of light polypeptide chain, all linked together by disulphide bonds. They have a binding domain for another molecule such that a given immunoglobulin binds specifically to that another molecule.

The protein can be natural or partly or wholly synthetically produced. The term also covers any protein having a binding domain which is homologous to an immunoglobulin binding domain.

Herein the numbering of immunoglobulin amino acid residues is according to Kabat E.A. et al. in "Sequences of Proteins of Immunological Interest" U.S. Department of Health and Human Services 1987.

Antibody

This describes an immunoglobulin whether natural or partly or wholly synthetically produced. The term also covers any protein having a binding domain which is homologous to an immunoglobulin binding domain. These proteins can be derived from natural sources, or partly

or wholly synthetically produced.

Example antibodies are the immunoglobulin isotypes and the Fab, $F(ab^1)_2$, scFv, Fv, dAb, Fd fragments.

Immunoglobulin Superfamily

This describes a family of polypeptides, the members of which have at least one domain with a structure related to that of the variable or constant domain of immunoglobulin molecules. The domain contains two β-sheets and usually a conserved disulphide bond (see A.F. Williams and A.N. Barclay 1988 Ann. Rev Immunol. 6 381-405.

Example members of an immunoglobulin superfamily are CD4, platelet derived growth factor receptor (PDGFR), intercellular adhesion molecule. (ICAM). Except where the context otherwise dictates, reference to immunoglobulins and immunoglobulin analogs in this application includes members of the immunoglobulin superfamily and analogs thereof.

Interface

This describes the region on a given heavy or light chain of an immunoglobulin which associates with the complementary heavy or light chain.

Framework

Each chain of an immunoglobulin has a constant (C) and a variable (V) region. Each V region is made up from three complementarity determining regions (CDR) separated by four framework regions (FR). The CDRs are variable stretches of amino acid sequences and provide the

function of binding to another molecule. It is the possibility of variability which provides immunoglobulins with various binding specificities. The FRs are substantially constant stretches of amino acid sequences which interpose the CDRs.

In order that the present invention is more fully understood, it will now be described in more detail, firstly in general outline and secondly with reference to specific examples provided by way of illustration only and not by way of limitation. The following description refers to the figures in which:

Figure 1 shows the nucleotide and amino acid sequence of the VH domain of the anti-lysozyme antibody D1.3 cloned in pUC119;

Figure 2 shows mutant oligonucleotides for substitution of residues found in naturally occurring immunoglobulin heavy chains;

Figure 3 shows mutant oligonucleotides for substitution with homologous residues from Thy-1;

Figure 4 shows mutant sequences obtained by substitutions of Thy-1 residues into VHD1.3;

Figure 5 shows oligonucleotides for the random substitution of asparagine, serine or threonine into VHD1.3;

Figure 6 shows a graph illustrating lysozyme binding activity of TG1 (control), VHD1.3, VHTHY-1 and VHTHY-2;

Figure 7 shows a graph illustrating lysozyme binding activity of pUC119 (control), VHD1.3, VHMutTrp, VHMutLeu

and VHThy-3;

Figure 8 shows a graph illustrating lysozyme binding activity of pUC119 (control), VHD1.3, VHThy-1, Thy-2, VHThy-1, Thy-3 and VHMutWD;

Figure 9 shows the nucleotide sequence of the vector fdPs/Bs around the cloning site and indicates the Pst1 and BstE11 restriction sites; and

Figure 10 shows a graph illustrating lysozyme binding activity of vector fdPs/Bs (control), fdVHThy1Thy2, fdVHD1.3 and phage antibody D1.3.

The applicant has devised three related strategies for the selection of alterations to the frameworks. This invention enables the generation of antibodies and single variable domains eg. VH domains with improved properties for <u>in vivo</u> and <u>in vitro</u> use.

1. <u>Substitution with VH Interface Residues Found in Naturally Occurring Immunoglobulin Heavy Chains</u>

There is considerable amino acid sequence homology between different immunoglobulins. Homologies are detected by lining up different sequences one above the other, and sliding the chains along relative to one another, until the best level of identity between the different sequences is arrived at. These analyses are generally performed on a computer. As noted above, the framework residues are highly conserved, ie. particular amino acids will be present in the same positions in a series of different antibodies. Rare substitutions do occur, however, and the applicant searched for naturally-

occurring substitutions of VH interface residues. This was performed by reference to readily available compilations of antibody sequences [e.g. Kabat, E.A. et al. in "Sequences of Proteins of Immunological Interest" U.S. Department of Health and Human Services (1987)]. This analysis enables natural variants at any position to be identified. As these variant residues occur naturally (though infrequently) in antibodies, the applicant realised that they are less likely to severely disrupt domain structure. These natural substitutions most often occur once per antibody molecule. However, the applicant also combines the available substitutions from several antibodies, together in the same molecule. It is possible that different combinations of natural substitutions can be used.

2. <u>Substitution with Residues that have been Identified at the Homologous Position in Other Proteins of the Immunoglobulin Superfamily</u>

Of particular interest, are proteins which have domains containing the immunoglobulin fold (a tertiary structure characteristic of antibody domains that has been found in other proteins), but which do not associate with another domain.

Examples of molecules which contain a single domain homologous to immunoglobulin variable domains include: Thyl, Po myelin, CD7, CD28 and CTLA-4 [Williams, A.F. and Barclay, A.N. Ann. Rev. Immunol. 6, 381-405 (1988)].

Other proteins contain more than one unpaired,

antibody-like domain. CD4 and MRC OX-2 each contain N-terminal domains homologous to variable domains of antibody molecules ('V'-type domains, one in MRC OX2, two in CD4), and a C-terminal domain homologous to constant domains of antibody molecules ('C'-type domain, one each in MRC OX2 and CD4) [Williams, A.F. and Barclay, A.N. Ann. Rev. Immunol. 6, 381-405 (1988)].

The similarity in tertiary structure between these unpaired domain proteins and the domains comprising antibodies is reflected to some extent in homologies at the amino acid sequence level. Amino acid homologies are assigned in much the same way as in 1 above, although the alignment of sequences in this case can be problematic. For Thy-1, the alignment with VH residues 37,39,91 and 93 is relatively straightforward. However, VH residues 45 and 47 have been published in two different alignments with Thy-1 [A.F. Williams and J. Gagnon, Science 216 696-703, 1982; A.F. Williams and A.N. Barclay, Ann. Rev. Immunol. 6 381-405, 1988]. Exploration of alternative substitutions may be necessary to identify the most appropriate, in cases where there is more than one possible alignment.

There is a possibility that the modified VH interface exposed in the VH single domain would be antigenic in humans. This would be a disadvantage for *in vivo* therapeutic use. Substitution with residues which occur naturally in human Thy-1 may reduce this potential antigenic response.

3. Semi-Random Insertion of Polar Amino Acids at Interface Residues

This strategy enables the improvement of single domains using a less direct method. Oligonucleotides for mutagenesis are synthesized with a mixture of bases in some positions to give ambiguities for amino acid insertion at certain triplets. For example, the present applicant has designed a strategy which will allow the random insertion of the highly polar residues asparagine, serine or threonine by the use of codons with ambiguities at the second position. These residues could be inserted for instance at the interface positions 37,45,47,93 and 103 of the VH domain. The 243 possible frameworks resulting from this mutagenesis can then be screened to identify which of the semi-random combinations have the desired properties. Strategies for screening the resulting semi-random population would include estimation of antigen binding affinity and non-specific binding by ELISA (see below).

This is one example of a number of strategies which can be used to vary interface residues in a semi-random manner without recourse to the strategies outlined in 1) and 2) above.

The present invention allows the binding affinity and specificity of identified antibodies to be incorporated into single domain molecules with improved properties for *in vivo* and *in vitro* applications.

The frameworks generated using the model system

using the VH domain of antibody D1.3 (VHD1.3) can be used as frameworks for antibodies of any specificity by replacement of part(s), or all, of its CDRs with those from an antibody molecule of desired affinity and specificity. There are numerous ways in which this can be accomplished. For instance, following the determination of the sequence of the CDRs of the antibody of desired properties (eg., binding specificity), an oligonucleotide or series of oligonucleotides is synthesized which encodes these CDRs and the framework regions containing the nucleotide substitutions necessary to make the antibody/antibody domain more polar. This oligonucleotide can then be amplified using PCR, cloned into a suitable vector such as pUC119, and the product expressed in bacteria.

Alternatively, the relevant changes at the interface can be introduced into an existing single domain antibody of desired specificity to improve the properties of that antibody. This can be achieved by a variety of methods, for example by site-directed mutagenesis, or PCR (for example using the method of Hemsley, A. et al. [Nuc. Acids Res. 16. 6545-6551 (1989)].

In addition to this, the improved single domain antibodies show reduced non-specific. The properties of the single chain variable domains provided hereby enable selection of single domain antibodies of the desired specificity and affinity, for example, using single domains cloned into fd phage. The frameworks described

in this invention could be cloned into fd phage and the existing CDRs replaced by repertoires of CDRs to create a new population of single domain antibody molecules that can be screened for the desired binding specificities. Alternatively, for instance, single domain antibodies may be isolated as described by Ward et al (1989, supra), but may require improvements in their affinity and specificity. The CDRs from these antibodies can be cloned into the polar frameworks described in this invention and inserted into single domains cloned into fd phage. Random mutagenesis of these CDRs can then be performed and selection of antibodies of the desired affinity and specificity performed using affinity methods.

Examples 15 and 16 herein show that a derivative of VHD1.3 with a more polar framework can be displayed on phage as a fusion with gene III protein (McCafferty J. et al. 1990 Nature 348 p552-554) with retention of binding activity. Display on phage would enable the generation of combinations of substitutions at framework residues by semi-random mutagenesis procedures (an example of which is given in example 11) and the subsequent selection of those with favourable binding properties.

Since the improved single domain antibodies of the present invention constitute superior versions of conventional single domain antibodies, they can be used in many of the ways as are immunoglobulins (Ig) and their superfamily of molecules or fragments. For example, Ig

molecules have been used in research, therapy (e.g. cancer therapy, modulation of immune status and therapy of diseases caused by pathogens), diagnosis (for example, estimation of hormonal status), in modulation of activities of hormones or growth factors, in detection, in biosensors, in catalysis, in purification of other molecules and in screening regimes for therapeutic compounds in the pharmaceutical industries. The lower non-specific binding of the improved single domain antibodies should prove especially useful for the above applications.

The increased hydrophilicity may be of particular importance for their use as binding molecules in affinity chromatography, especially weak affinity chromatography (Zopf, D. and Ohlson, S., Nature 346 87-89, 1990).

Anti-idiotypic improved single domain antibodies can also be made. Anti-idiotypic specificities [Methods Enzymol. 178. J.J. Langone ed. Academic Press (1989)] are made in a two-stage process. Firstly, antibody A directed against a particular antigen or epitope is itself used to raise other antibodies. A proportion of the anti-A antibodies, antibodies B, will be directed against the antigen combining sites of antibody A, such that the antigen combining sites of B are complementary to that of A. In effect, the antigen combining site of antibody B, the anti-idiotype, mimics in structure the original antigen or epitope recognised by antibody A. The original antigen can be a protein or any other

compound, for example a carbohydrate or a steroid, and the antibody used at any stage in the procedure could be an improved single domain antibody. The final anti-idiotypic antibody can be an improved single domain antibody produced as described herein, or a molecule of the immunoglobulin superfamily from which the anti-idiotypic determinant(s) are transferred into an improved single domain antibody framework.

Such anti-idiotypic molecules are advantageous in a variety of applications [Methods Enzymol. 178. J.J. Langone ed. Academic Press (1989)]. These include vaccines for treating cancers and diseases caused by bacteria, viruses and parasites. They may be used for blocking cellular receptors for the aforementioned pathogens as well as blocking cellular receptors for hormones. They may also be advantageous in diagnostic procedures, for example in place of antigen or peptide in ELISA. Anti-idiotypic specificities are known to be useful in the pharmaceutical industries [Methods Enzymol. 178. J.J. Langone ed. Academic Press (1989)].

The present invention relates to improved single domain antibodies and receptors derived from molecules of the immunoglobulin (Ig) superfamily, methods for selecting and effecting said improvements, and methods for and kits use of said antibodies or receptors in research, therapy, diagnosis, purifications, catalysis and discovery of novel therapeutics.

EXAMPLE 1. PREPARATION OF VH D1.3 WITH A MORE POLAR

FRAMEWORK BY SUBSTITUTION OF VALINE 37 WITH GLUTAMINE

Figure 1 shows the nucleotide and amino acid sequence of the pUC119 VHD1.3 clone used for mutagenesis studies (Ward et al, 1989, supra). The amino acid residues in the VH domain that interact with the VL have been identified as 37,39,45,47,91,93 and 103 (Amit et al, (1986) supra, Chothia, C. et al (1986) supra). Amino acids occurring naturally in VH domains were surveyed using a compilation of immunoglobulin sequences [for example, Kabat, E.A. et al. in "Sequences of Proteins of Immunological Interest" U.S. Department of Health and Human Services (1987)]. Substitutions have been found at positions 37,39,45,47,91,93 and 103 in naturally occurring heavy chains. The most polar substitutions at each of these residues in antibodies sequenced to date were chosen for replacement of interface residues by mutagenesis in the following examples.

Residue 37 is valine in 385 out of 434 sequences surveyed. In 48 other sequences, an aliphatic amino acid was substituted. In the single remaining example, a case of human heavy chain disease, a significantly more polar residue, glutamine, was substituted. An oligonucleotide was designed for incorporation of this change in VHD1.3 (VHMUTVAL, shown in figure 2). This mutation and the other mutations described in the examples below can be assembled in different combinations to make further novel derivatives.

In Vitro Mutagenesis

(1) The oligonucleotides detailed in fig. 2 were synthesised on an Applied Biosystems 391 DNA synthesiser and purified on Urea-acrylamide gels using standard techniques [Sambrook, J. et al. in "Molecular Cloning: a laboratory manual (second edition)". Cold Spring Harbor Laboratory Press, 11.23] prior to in vitro mutagenesis.

(2) Preparation of single stranded DNA template for mutagenesis. The VH D1.3 antibody gene (Ward et al, 1989, supra) used to exemplify this invention was carried on plasmid pUC 119 (Sambrook, J. et al. in "Molecular Cloning: a laboratory manual (second edition)". Cold Spring Harbor Laboratory Press 1.14). Single-stranded template DNA was prepared by infecting TG1 cells carrying the plasmid with M13 KO7 helper phage using standard techniques for growth and purification (Sambrook, J. et al. in "Molecular Cloning: a laboratory manual (second edition)". Cold Spring Harbor Laboratory Press 4.46).

(3) Site-directed mutageneses were performed using the "In vitro Mutagenesis system, Oligonucleotide directed (version 2)" (Amersham International) exactly as per manufacturers instructions. Ampicillin-resistant colonies resulting from the mutageneses were grown overnight in 2YT (2YT=per litre of water, 16g Bacto-tryptone, 10g Yeast extract, 5g NaCl) containing 100µg/ml ampicillin. These cultures were diluted in fresh 2YT and single stranded template DNA prepared by M13 KO7 infection as in (2) above. Mutants were verified by DNA sequencing using the Sequenase version 2.0 kit.

EXAMPLE 2 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF GLUTAMINE 39 WITH GLUTAMATE

Using similar logic and methods to those described in example 1, an oligonucleotide VHMUTGLN (fig. 2) was designed for the introduction of a glutamate residue at position 39. This substitution is found in 1 of 420 heavy chains surveyed. Glutamate is considered to be marginally more polar than glutamine [Rose et al, Science 229 834-838, 1985].

EXAMPLE 3 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF LEUCINE 45 WITH GLUTAMINE

Using similar logic and methods to those described in example 1, an oligonucleotide VHMUTLEU (figure 2) was designed which introduces a substitution of leucine with glutamine at residue 45 (fig. 2). This modification is found in 2 of 402 sequences surveyed (396 have leucine at this position). One of the antibodies containing glutamine at this position is a mouse antibody specific for anti-B1, 6D-galactan.

EXAMPLE 4 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF TRYPTOPHAN 47 WITH ASPARTATE

Using similar logic and methods to those described in example 1, an oligonucleotide VHMUTWD (figure 2) was designed for the introduction of aspartate at position 47. This substitution is found in one of 392 heavy chains surveyed.

EXAMPLE 5 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF TYROSINE 91 WITH THREONINE

Using similar logic and methods to those described in example 1, an oligonucleotide VHMUTTYR was designed for the introduction of threonine at position 91. This substitution is found in one of 398 heavy chains surveyed.

EXAMPLE 6 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF ALANINE 93 WITH SERINE

Using similar logic and methods to those described in example 1, an oligonucleotide VHMUTALA (figure 2) was designed for substitution of alanine 93 with serine. This substitution is found in 4 of 410 heavy chains surveyed. One of these is in a mouse anti-B2,1 fructosan.

EXAMPLE 7 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF TRYPTOPHAN 103 WITH GLUTAMATE OR TYROSINE

Using similar logic and methods to those described in example 1, two oligonucleotides VHMUTTRP and VHMUTWY (figure 2) were designed for the introduction of glutamate and tyrosine respectively at position 103. These substitutions are found once each in 308 heavy chains surveyed. Glutamate is much more polar than tryptophan. Tyrosine, although more polar than tryptophan is a more conservative substitution.

EXAMPLE 8 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF RESIDUES IN POSITIONS 37,39 AND 47 WITH HOMOLOGOUS RESIDUES FROM THE IMMUNOGLOBULIN FAMILY PROTEIN THY-1

Thy-1 is a single domain protein of the immunoglobulin superfamily. Alignment of the residues of Thy-1 with those of immunoglobulin heavy chains has been performed by Williams (A.F. Williams and J. Gagnon Science 216 696-703 1982; A.F. Williams and A.N. Barclay Ann. Rev. Immunol. 6 381-405, 1988). Although residues 37,39,91 and 93 of the VH domain interface were aligned with the same residues of Thy-1 in both publications, the residues at positions 45 and 47 were aligned with different residues reflecting the lower degree of homology of adjacent amino acids from these positions. The oligonucleotide VHTHY-1 (figure 3) was designed to incorporate the most polar residues at positions 37,39 and 47 found at homologous Thy-1 positions using the alignment published by Williams & Gagnon (1982), supra). Mutagenesis was performed as in example 1. The amino acid substitutions generated are shown in figure 4.

<u>EXAMPLE 9 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF RESIDUES IN POSITIONS 91,93 AND 103 WITH HOMOLOGOUS RESIDUES FROM THE IMMUNOGLOBULIN FAMILY PROTEIN THY-1</u>

Using the strategy described in example 8, an oligonucleotide VHTHY-2 (figure 3) was designed to incorporate the most polar substitutions of VH residues 91,93 and 103 at homologous Thy-1 residues [Kabat, E.A. et al. in "Sequences of Proteins of Immunological Interest" U.S. Department of Health and Human Services (1987). Another oligonucleotide VHTHY-3 (figure 3) was

designed to incorporate the residues found at these positions in rat brain Thy-1 which appears most polar overall at residues homologous to the VH interface. Mutagenesis is as described in example 1. The amino acid changes generated are shown in figure 4.

EXAMPLE 10 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF RESIDUES IN POSITIONS 37,39,47,91,93 AND 103 WITH HOMOLOGOUS RESIDUES FROM THE IMMUNOGLOBULIN FAMILY PROTEIN THY-1

The amino acid changes detailed in examples 8 and 9 were combined by performing a mutagenesis experiment as in example 1 using the DNA sequence of mutant protein VHTHY-1 as template and mutant oligonucleotides VHTHY-2 and VHTHY-3 (figure 3) to incorporate site directed changes.

EXAMPLE 11 PREPARATION OF VH D1.3 WITH A MORE POLAR FRAMEWORK BY SUBSTITUTION OF RESIDUES IN POSITIONS 37,45,47,93 AND 103 WITH ASPARAGINE, THREONINE OR SERINE

Oligonucleotides are prepared containing the triplet GXT at each of the positions 37,39,45,47,91,93 and 103 (where X is a random mixture of the bases C,G and T; figure 5). Use of these as mutagenesis primers as in example 1 would generate the insertion of Ser, Thr and Asn respectively depending on which base was incorporated. Derivatives generated are then screened for antigen binding and improved properties.

EXAMPLE 12: ASSESSMENT OF ANTIGEN BINDING STATUS OF VHTHY-1 AND VHTHY-2 MUTANT SINGLE DOMAIN ANTIBODIES

VH D1.3 interface mutants VHThy-1 and VHThy-2 constructed as described in example 8 and 9 were assessed for lysozyme binding activity. Antigen binding status of mutant single domain antibodies was determined by ELISA (Enzyme Linked Immuno adSorbent Assay) according to techniques well known in the art.

This is just one of a whole range of methods that can be used to measure antigen-antibody binding. Others include Western blotting, competitive radioimmunassay and fluorescence quench.

The ELISA for lysozyme binding by mutant single domains was undertaken as follows:

Overnight cultures of ampicillin-resistant clones were diluted 1 in 10 into fresh 2YT (with 100μg/ml ampicillin) and grown for 1hr at 37°C. Isopropyl β-D-thiogalactopyranoside (IPTG) was added to 1mM final concentration and the cells cultured for a further 24-30 hrs at 37°C. Supernatants were prepared by centrifugation for use directly in the ELISA.

1) The plates (Falcon microtest III flexible plate) were coated with 200μl per well of 1mg/ml hen egg lysozyme in 50mM NaHCO3, pH 9.6 overnight at room temperature.

2) The wells were rinsed with three washes of phosphate buffered saline (PBS), and blocked with 300μl per well 2% skimmed milk powder in PBS for two hours at 37°C.

3) The wells were rinsed with three washes of PBS and 200μl of culture supernatant were added and incubated for

two hours at room temperature.

4) The wells were washed three times with 0.05% Tween 20 in PBS and three times in PBS.

5) 200μl of a suitable dilution (1 in 1000) of rabbit polyclonal antiserum against the Fv fragment in 2% skimmed milk powder in PBS was added to each well and incubated for two hours.

6) Washes were repeated as in (4).

7) 200μl of a suitable dilution (1 in 5000) of goat anti-rabbit antibody (Sigma) coupled to horseradish peroxidase, in 2% skimmed milk powder in PBS, was added to each well and incubated at room temperature for one hour.

8) Washes were repeated as in (4).

9) 200μl 2,2' azino-bis (3-ethylthiazolinesulphonic acid) [Sigma] (0.55mg/ml in citrate buffer (citrate buffer comprises: 54mls 50mM citric acid plus 46mls 50mM trisodium citrate per 100mls), with 1μl 30% hydrogen peroxide:water per 10 mls) was added to each well and the colour allowed to develop for up to ten minutes at room temperature.

The reaction was stopped by adding 0.05% sodium azide in 50mM citric acid pH4.3. ELISA plates were read at 405nm in a Titertek Multiskan M.C. to give an optical density reading for each well. The optical density reading is in proportion to the amount and affinity of the primary antibody used in the ELISA, in this case a single domain antibody.

The result shown in figure 6 demonstrates that these mutants had retained their ability to bind lysozyme. The VHThy-1 mutant appears to have higher affinity/quantity of lysozyme binding activity than the parent VH D1.3, whereas mutant VHThy-2 has slightly less.

Given the complexity of the problem (see above), this result was surprising and reinforces the point that the properties of VH single domain antibodies can surprisingly be improved without unduly compromising antigen binding, as taught in this document.

<u>EXAMPLE 13 ASSESSMENT OF ANTIGEN BINDING STATUS OF VHMUTTRP, VHMUTLEU AND VHTHY-3 MUTANT SINGLE DOMAIN ANTIBODIES</u>

VHD1.3 interface mutants VHMutLeu, VHMutTrp and VHThy-3 constructed as described in examples 3,7 and 9 respectively were assessed for lysozyme binding activity., Antigen binding status of mutant single binding domains was determined by ELISA as described in example 12. The result shown in figure 7 demonstrates that these mutants had retained their ability to bind lysozyme.

<u>EXAMPLE 14 ASSESSMENT OF ANTIGEN BINDING STATUS OF VHTHY-1, THY-2; VHTHY-1, THY-3 AND VHMUTWD MUTANT SINGLE DOMAIN ANTIBODIES</u>

VHD1.3 interface mutants VHThy-1, Thy-2; VHThy-1, Thy-3 and VHMutWD constructed as described in examples 10 and 4 respectively were assessed for lysozyme binding activity. Antigen binding status of mutant single

binding domains was determined by ELISA as described in example 12. The result shown in figure 8 demonstrates that these mutants had retained their ability to bind lysozyme.

Thus, extensive changes (6 amino acid substitutions) can be made at the VH interface increasing the polarity of the domain without affecting the ability to bind lysozyme. The approach used to select substitutions taught in this document may be expected to be applicable to any VH domain.

EXAMPLE 15 CLONING OF GENE ENCODING VHTHY-1, THY-2 INTO fdPs/Bs FOR DISPLAY ON PHAGE

The gene encoding the derivative VHThy-1, Thy-2, generated in example 10 by using the DNA sequence of the mutant protein VHThy-1 as template and the mutant oligonucleotide VHThy-2 to incorporate site directed changes, was subcloned into the vector fdPs/Bs for display of this VH domain on phage as a fusion with gene III protein. The vector fdPs/Bs is similar to fdCAT1 (McCafferty, J. et al, 1990 Nature 348 p552-554) except that it contains Pst1 and BstB11 restriction sites for cloning (Figure 9).

A miniprep of pUC119 VHThy-1, Thy-2 DNA was prepared using standard procedures (Sambrook et al, 1989 supra). The VHThy-1, Thy-2 encoding sequences were amplified by PCR using the primers RVHTHYFOR and KSJ6.

RVHTHYFOR

5' TGA GGA GAC GGT GAC CGT GGT GCC TTG GCC AGT G 3'

This incorporates a BstE11 site at the 3' end of the VHThy-1, Thy-2 gene.

KSJ6

5' AGG TGC AGC TGC AGG AGT CAG G 3'

This incorporates a PstI site at the 5' end of the VHThy-1, Thy-2 gene.

PCR was performed using 20mM Tris (pH7.3 at 70°C), 50mM KCl, 4mM $MgCl_2$, 0.01% gelatin with 10µM each oligonucleotide, 1mM each dNTP, 5 units Taq polymerase and approximately 50ng pUC119VHThy-1, Thy-2 DNA in a total volume of 100µl. The product of the PCR reaction was ethanol precipitated and resuspended in 20µl 10mM Tris, pH8.0, 0.1 mM EDTA, A 10µl portion was digested using PstI (20 units) and BstE11 (20 units) in NEB buffer 2 in a total volume of 50µl at 37°C for 2h (restriction enzymes obtained from New England Biolabs, CP Labs, Bishops Stortford). Following digestion, the reaction mixture was phenol extracted and ethanol precipitated. The product was electrophoresed on a 1% agarose Tris-acetate-EDTA gel and the band of approximately 350bp excised and the DNA purified using Geneclean (Bio 101, La Jolla, California). Vector DNA (fdPs/Bs RF form) was prepared using standard procedures (Sambrook et al, 1989 supra). This DNA (1.2µg) was digested with PstI and BstE11 (50 units) in 100µl NEB buffer 3 at 37°C for 90 min. The products were phenol extracted and ethanol precipitated and the resuspended DNA was phosphatased as described by Sambrook et al, (1989 supra). A preparative

0.7% Tris-borate-EDTA agarose gel was performed and the band of approximately 9kb excised and the DNA purified using Geneclean and resuspended in 10 µl 10mM Tris, pH8.0, 0.1mM EDTA. Ligation was performed using 5µl each of the digested vector and insert DNA using 200 units of T4 DNA ligase in 10µl NEB ligase buffer. The ligation mixture (8µl) was transformed into competent E.coli MC1061 cells prepared according to Sambrook et al, 1989 supra) and the mixture plated on 2YT agar containing 20µg/ml tetracycline. Colonies were picked, single stranded DNA was prepared (Sambrook et al, 1989 supra) and the DNA was sequenced using a Sequenase 2.0 kit (United Sates Biochemical, Cleveland, U.S.A.). The sequence of the insert corresponded to VHThy-1, Thy-2. The derivative has been named fdVHThy-1, Thy-2.

A clone of VHD1.3 in fdPs/Bs was prepared starting from pUC119VHD1.3. The insert encoding VHD1.3 was prepared by digestion of pSW1-VHD1.3-TAG1 (Ward E.S. et al., 1989 supra) with PstI and BstE11. Other procedures were as above. This derivative has been named fdVHD1.3.

EXAMPLE 16 ELISA ASSAY OF VHTHY-1, THY-2 DOMAIN DISPLAYED ON PHAGE

The fdVHThy-1, Thy-2 phage constructed in example 15 was shown to be functional in the binding of the antigen, lysozyme, using an ELISA assay.

Viral particles were prepared by growing E.coli MC1061 cells containing fdVHThy-1, Thy-2; fdVHD1.3; phage antibody D1.3 (displaying scFvD1.3; McCafferty, J. et al,

1990 Nature 348 p552-554) or fdPs/Bs in 50ml 2YT medium containing 15µg/ml tetracycline for 16 to 24h. The culture supernatant was collected by centrifugation for 10 min at 10000rpm in an 8 x 50ml rotor. Phage particles were precipitated by adding 1/5 volume 20% polyethylene glycol (PEG)/2.5M NaCl and leaving at 4°C for 1h. Phage particles were pelleted by centrifugation for 15 min as described above and the pellets resuspended in sterile 10mM Tris, pH8.0 1mM EDTA containing 1% gelatin to 1/40 the of the original volume.

1. ELISA plates were coated with lysozyme and blocked with PBS containing skimmed milk powder as described in example 12.

2. Wells were rinsed with PBS.

3. Concentrated phage (200µl) was added to each well as appropriate and incubated at room temperature for 2h.

4. The wells were washed three times with 0.5% Tween 20 in PBS and three times with PBS.

5. Sheep anti-M13 serum (200µl; 1 in 1000) in PBS containing 2% skimmed milk powder was added to each well and incubated for 1h.

6. Washing was repeated as in 4.

7. Peroxidase conjugated rabbit anti-goat immunoglobulin (200µl; 1 in 5000; Sigma) was added and incubated for 1h.

8. Washes were repeated as in 4.

9. Peroxidase substrate was added as in example 12 and colour allowed to develop for 1h.

Both fdVHThy-1, Thy-2 and fdVHD1.3 gave ELISA signals 4 to 5 times above the value obtained with fdPs/Bs, whereas with phage antibody D1.3, the signal was approximately 8 times that with fdPs/Bs (Figure 10). Thus modification of the VH interface residues does not affect the ability of the domain to bind lysozyme when displayed on phage.

It will be appreciated by those skilled in the art that the present invention has been described above by way of example only, and that considerable modifications to the procedure may be made to effect a similar outcome without departing from the scope of the invention.

CLAIMS

1. A single chain variable domain, which is a synthetic analogue of another single chain variable domain of a member of an immunoglobulin family or superfamily, and in which analogue, one or more interface amino acid residues of the domain is altered as compared to the said another domain, in which a said altered amino acid is substituted with a residue which occurs in an analogous position in a member of an immunoglobulin family or superfamily, so that the analogue is more hydrophilic than the said another domain.

2. A single chain variable domain according to claim 1 in which a said altered amino acid residue is in a framework region.

3. A single chain variable domain according to claim 1 or claim 2 in which a said altered amino acid residue is in a complementarity determining region.

4. A single chain variable domain according to any one of claims 1 to 3 wherein the synthetic analogue has essentially the same binding activity as the said another domain.

5. A single chain variable domain according to claim 2 or claim 3 in which the amino acid sequence of a complementarity determining region is additionally altered by way of amino acid substitution, deletion,

addition, or inversion, to alter the specificity and/or binding characteristics of the analogue as compared to the natural domain.

6. A single chain variable domain according to any one of claims 1 to 5, which is a synthetic analogue of a single variable immunoglobulin heavy chain domain.

7. A single chain variable domain according to claim 6, in which one or more of the amino acid residues 37, 39, 45, 47, 91, 93 and 103 is altered.

8. A single chain variable domain according to claim 6 or claim 7 in which the amino acid alterations comprise one or more of the following:

i) substitution of valine 37 with glutamine or threonine;

ii) substitution of glutamine 39 with glutamate;

iii) substitution of leucine 45 with glutamine;

iv) substitution of tryptophan 47 with aspartate or glycine;

v) substitution of tyrosine 91 with threonine, serine or methionine;

vi) substitution of alanine 93 with serine or glutamate;

vii) substitution of tryptophan 103 with glutamate tyrosine or threonine;

viii) substitution of valine 37, leucine 45, tryptophan 47, alanine 93 and/or tryptophan 103

with any of asparagine, threonine or serine;

ix) substitution of valine 37 with threonine and glutamine 39 with glutamate and tryptophan 47 with glycine;

x) substitution of tyrosine 91 with serine or methionine and alanine 93 with glutamate and tryptophan 103 with threonine.

9. A single chain variable domain according to any one of claims 1 to 8 which is coupled to a further molecular moiety.

10. An immunoglobulin single chain variable domain according to claim 9 wherein the further molecular moiety is an enzymic-, fluorescent-, or radio-label, or a portion of an immunoglobulin.

11. A diagnostic kit which comprises a single chain variable domain according to any one of claims 1 to 10, together with one or more ancillary reagents for carrying out the diagnostic test.

12. A therapeutic composition which comprises at least a single chain variable domain according to any one of claims 1 to 10.

13. A method for making a single chain variable domain which is a hydrophilic synthetic analogue of another single chain variable domain of a member of an immunoglobulin family or superfamily, which comprises:

(i) inspecting the interface regions of a said single

chain variable domain to identify hydrophobic amino acid residues; and

(ii) producing a said analogue of said single chain variable domain in (i) in which one or more of said hydrophobic residues is substituted with a less hydrophobic residue which occurs in an analogous position in a member of an immunoglobulin family or superfamily.

14. A method according to claim 13 which comprises:

(a) obtaining the nucleotide sequence encoding one or more of the identified hydrophobic amino acid residues;

(b) using site directed mutagenesis to alter the nucleotide sequence to introduce a triplet coding for the substitute amino acid,

(c) using the altered nucleotide sequence in a recombinant expression system to express the synthetic analogue.

15. A method according to claim 13 or claim 14 wherein more than one amino acid residue is substituted.

16. A method according to any one of claims 13 to 15 wherein the substitute amino acids are derived from naturally monomeric members of the immunoglobulin superfamily.

17. A method according to claim 16 wherein the naturally monomeric member is Thy-1.

18. A method according to any one of claims 13 to

17 wherein the synthetic analogue has essentially the same binding activity as the said another domain.

```
                       rbs              M  K  Y  L  L  P  T  A  A
GCATGCAAATTCTATTTCAAGGAGACAGTCATAATGAAATACCTATTGCCTACGGCAGCC
        10        20        30        40        50        60

        PelB   leader
 A  G  L  L  L  L  A  A  Q  P  A  M  A  Q  V  Q  L  Q  E  S
GCTGGATTGTTATTACTCGCTGCCCAACCAGCGATGGCCCAGGTGCAGCTGCAGGAGTCA
        70        80        90       100       110       120
                                                  PstI

 G  P  G  L  V  A  P  S  Q  S  L  S  I  T  C  T  V  S  G  F
GGACCTGGCCTGGTGGCGCCCTCACAGAGCCTGTCCATCACATGCACCGTCTCAGGGTTC
       130       140       150       160       170       180

 S  L  T  G  Y  G  V  N  W  V  R  Q  P  P  G  K  G  L  E  W
TCATTAACCGGCTATGGTGTAAACTGGGTTCGCCAGCCTCCAGGAAAGGGTCTGGAGTGG
       190       200       210       220       230       240

                         VHD1.3
 L  G  M  I  W  G  D  G  N  T  D  Y  N  S  A  L  K  S  R  L
CTGGGAATGATTTGGGGTGATGGAAACACAGACTATAATTCAGCTCTCAAATCCAGACTG
       250       260       270       280       290       300

 S  I  S  K  D  N  S  K  S  Q  V  F  L  K  M  N  S  L  H  T
AGCATCAGCAAGGACAACTCCAAGAGCCAAGTTTTCTTAAAAATGAACAGTCTGCACACT
       310       320       330       340       350       360

 D  D  T  A  R  Y  Y  C  A  R  E  R  D  Y  R  L  D  Y  W  G
GATGACACAGCCAGGTACTACTGTGCCAGAGAGAGAGATTATAGGCTTGACTACTGGGGC
       370       380       390       400       410       420

 Q  G  T  T  L  T  V  S  S
CAAGGCACCACTCTCACAGTCTCCTCATAATAAGAGCTA
       430       440       450
```

Fig 1.

FIGURE 2.

Sequence of minus (-) strand of VHD1.3

	5' CA AAT CAT TCC CAG CCA CTC CAG ACC CTT TCC TGG AGG CTG GCG AAC CCA GTT TAC ACC A 3'
VHMUTWD (47 Trp->Asp)	5' CA AAT CAT TCC CAG ATC CTC CAG ACC CTT 3'
VHMUTLEU (45 Leu->Gln	5' AG CCA CTC CTG ACC CTT TC 3'
VHMUTGLN (39 Gln->Glu)	5' C TGG AGG CTC GCG AAC CCA 3'
VHMUTVAL (37 Val->Gln)	5' TGG AGG CTG GCG CTG CCA GTT TAC ACC A 3'

Mutagenizing Oligonucleotide	**Sequence of minus (-) strand of VHD1.3**
	5' GGT GCC TTG GCC CCA GTA GTC AAG CCT ATA ATC TCT CTC TCT GGC ACA GTA GTA CCT GGC TGT3'
VHMUTTRP (103 Trp->Glu)	5' GT GCC TTG GCC CTC GTA GTC AAG CCT A3'
VHMUTWY (103 Trp->Tyr)	5' GT GCC TTG GCC GTA GTA GTC AAG CCT A3'
VHMUTALA (93 Ala->Ser)	5' T CTC TCT GGA ACA GTA GTA 3'
VHMUTTYR (91 Tyr->Thr)	5' TC TCT GGC ACA GGT GTA CCT GGC TGT3'

FIGURE 3.

Mutagenising Oligonucleotide	**Sequence of minus (-) strand of VHD1.3**
	5' CA AAT CAT TCC CAG CCA CTC CAG ACC CTT TCC TGG AGG CTG GCG AAC CCA GTT TAC ACC A 3'
VHTHY-1	5' CA AAT CAT TCC CAG GCC CTC CAG ACC CTT TCC TGG AGG CTC GCG AGT CCA GTT TAC ACC A 3'

Mutagenising Oligonucleotide	**Sequence of minus (-) strand of VHD1.3**
	5'GGT GCC TTG GCC CCA GTA GTC AAG CCT ATA ATC TCT CTC TCT GGC ACA GTA GTA CCT GGC TGT3'
VHTHY-2 (most polar residues substituted)	5'GGT GCC TTG GCC AGT GTA GTC AAG CCT ATA ATC TCT CTC TCT TTC ACA GGA GTA CCT GGC TGT3'
VHTHY-3 (substitutions from rat brain Thy-1)	5'GGT GCC TTG GCC AGT GTA GTC AAG CCT ATA ATC TCT CTC TCT TTC ACA CAT GTA CCT GGC TGT3'

FIGURE 4

Mutant protein	Amino acid sequence of VHD1.3																		
	33 G GGT	34 V GTA	35 N AAC	36 W TGG	37 V GTT	38 R CGC	39 Q CAG	40 P CCT	41 P CCA	42 G GGA	43 K AAG	44 G GGT	45 L CTG	46 E GAG	47 W TGG	48 L CTG	49 G GGA	50 M ATG	51 I ATT
VHTHY-1					T ACT		E GAG								G GGC				

Mutant protein	Amino acid sequence of VHD1.3																				
	87 T ACA	88 A GCC	89 R AGG	90 Y TAC	91 Y TAC	92 C TGT	93 A GCC	94 R AGA	95 E GAG	96 R AGA	97 D GAT	98 Y TAT	99 R AGG	100 L CTT	101 D GAC	102 Y TAC	103 W TGG	104 G GGC	105 Q CAA	106 G GGC	107 T ACC
VHTHY-2					S TCC		E GAA										T ACT				
VHTHY-3					M ATG		E GAA										T ACT				

FIGURE 5

Oligo for insertion at residues 37, 45 and 47

5' CA AAT CAT TCC CAG GXT CTC GXT ACC CTT TCC TGG AGG CTG GCG GXT CCA GTT TAC ACC A3'

Oligo for insertion at residues 93 and 103

5' GT GCC TTG GCC GXT GTA GTC AAG CCT ATA ATC TCT CTC TCT GXT ACA GTA GTA CCT G3'

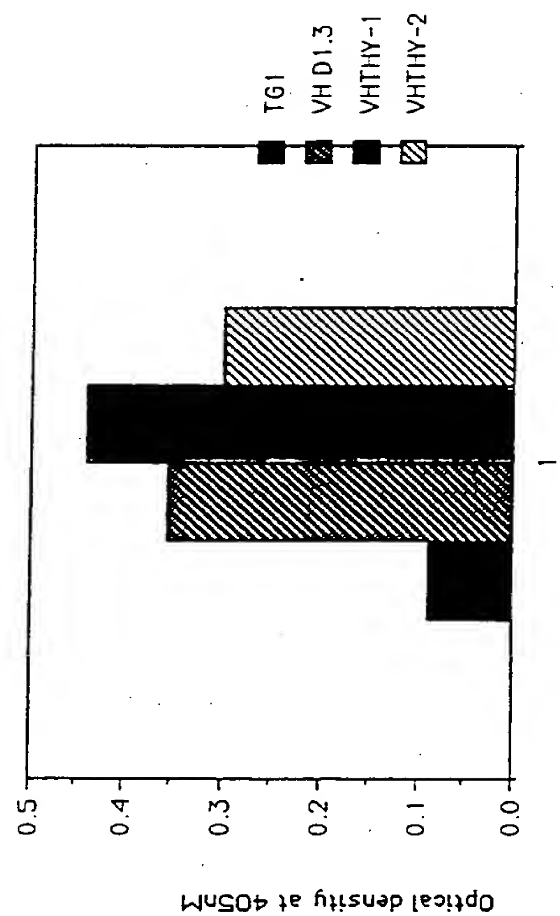
TG1
VH D1.3
VHTHY-1
VHTHY-2
0.5
0.4
0.3
0.2
0.1
0.0
1
Optical density at 405nM

Fig. 6.

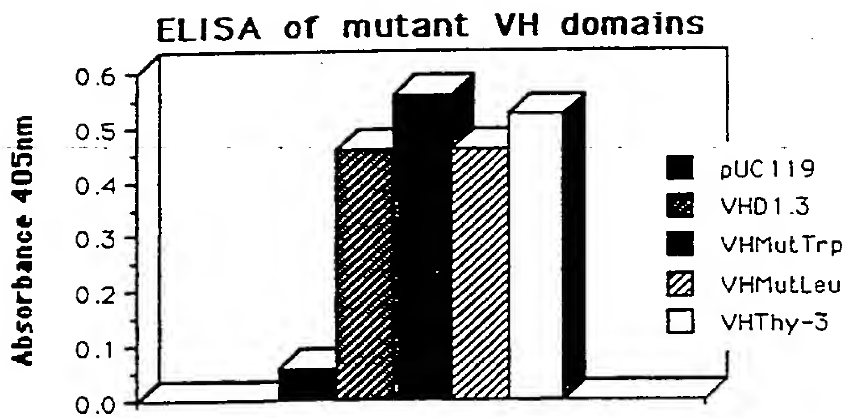
ELISA of mutant VH domains
Absorbance 405nm
0.6
0.5
0.4
0.3
0.2
0.1
0.0
pUC119
VHD1.3
VHMutTrp
VHMutLeu
VHThy-3

Figure 7.

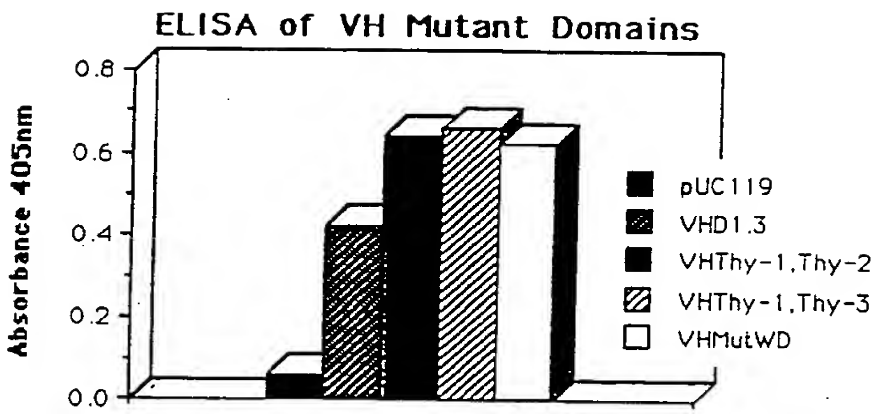
ELISA of VH Mutant Domains
Absorbance 405nm
0.8
0.6
0.4
0.2
0.0
pUC119
VHD1.3
VHThy-1,Thy-2
VHThy-1,Thy-3
VHMutWD

Figure 8.

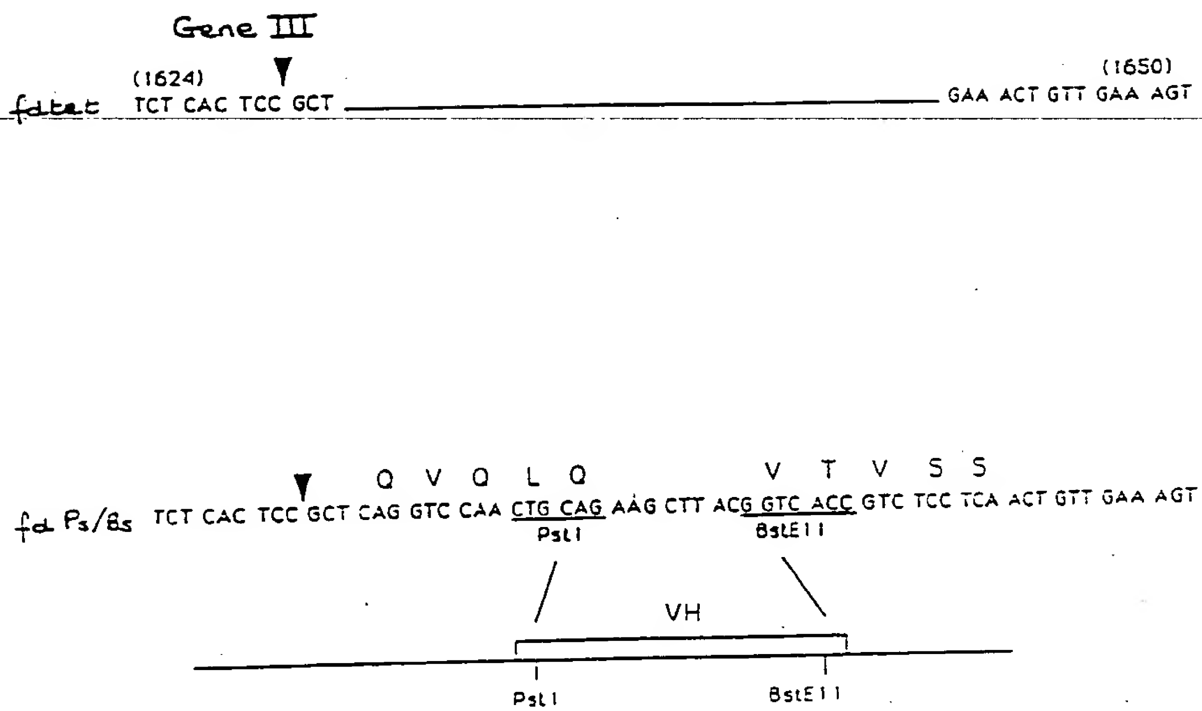
Gene III
(1624)
(1650)
fdtet TCT CAC TCC GCT
GAA ACT GTT GAA AGT
Q V Q L Q
V T V S S
fd Ps/Bs TCT CAC TCC GCT CAG GTC CAA CTG CAG AAG CTT ACG GTC ACC GTC TCC TCA ACT GTT GAA AGT
PstI
BstEII
VH
PstI
BstEII

Figure 9.

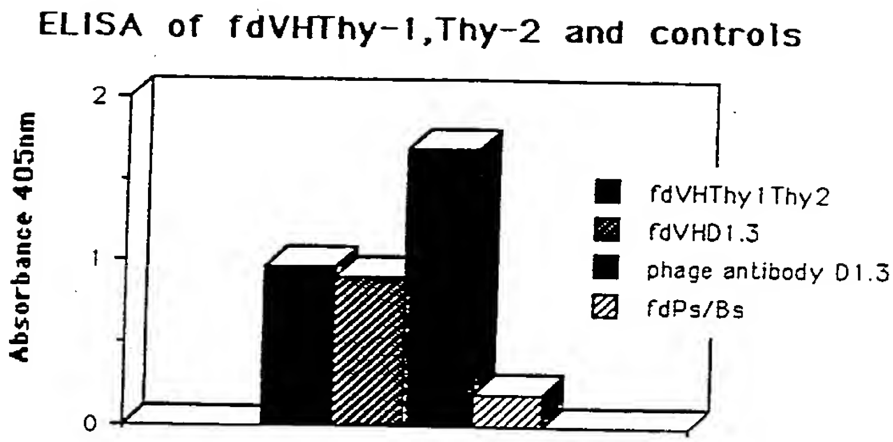
ELISA of fdVHThy-1,Thy-2 and controls
Absorbance 405nm
2
1
0
fdVHThy1Thy2
fdVHD1.3
phage antibody D1.3
fdPs/Bs

Figure 10.

INTERNATIONAL SEARCH REPORT

International Application No PCT/GB 91/01253

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/12; C12N15/13; C07K13/00; A61K37/02

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C07K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X Y	EP,A,0 368 684 (MEDICAL RESEARCH COUNCIL) 16 May 1990 See page 5, lines 19-32; page 6; page 7, lines 1-23; page 13, lines 3-30; examples 7-12; claims; figs. 10,12 ---	1-16,18 17
X Y	WO,A,8 809 344 (CREATIVE BIOMOLECULES, INC) 1 December 1988 See page 22, lines 5-32; pages 23,24,29; pages 30-57; claims ---	1-6, 9-16,18 17
Y	SCIENCE. vol. 216, 14 May 1982, LANCASTER, PA US pages 696 - 703; WILLIAMS, A.F. ET AL: 'Neuronal Cell Thy-1 glycoprotein: Homology with Immunoglobulin' cited in the application see the whole document ---	17

-/--

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
14 NOVEMBER 1991	22. 11. 91
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	NAUCHE S.A.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	NATURE. vol. 341, 12 October 1989, LONDON GB pages 544 - 546; WARD, E.S. ET AL.: 'Binding activities of a repertoire of single immunoglobulin variable domains secreted from Esherichia Coli.' cited in the application see the whole document	1-18
A	JOURNAL OF MOLECULAR BIOLOGY vol. 204, no. 1, 5 November 1988, ACADEMIC PRESS, H. B. JOVANICH, PUBL pages 155 - 164; JANIN, J. ET AL.: 'Surface, subunit interfaces and interior of oligomeric proteins.' see the whole document	1-18
A	PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA. vol. 82, July 1985, WASHINGTON US pages 4592 - 4596; NOVOTNY, J. ET AL.: 'Structural invariants of antigen binding : comparison of immunoglobulin VL-VH AND VL-VL domain dimers.'	1-18

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. GB 9101253 SA 49593

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on 14/11/91
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information.

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-0368684	16-05-90	AU-A- 4520189	28-05-90
		WO-A- 9005144	17-05-90
		JP-T- 3502801	27-06-91
WO-A-8809344	01-12-88	AU-B- 612370	11-07-91
		AU-A- 1804988	21-12-88
		EP-A- 0318554	07-06-89
		JP-T- 2500329	08-02-90

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82

XP 002046115 C07K3/18 C12P21/02F C12N15/66

A SHORT POLYPEPTIDE MARKER SEQUENCE USEFUL FOR RECOMBINANT PROTEIN IDENTIFICATION AND PURIFICATION

Pd. 10-1988 P.1204-1210 = (7)

Thomas P. Hopp, Kathryn S. Prickett, Virginia L. Price, Randell T. Libby[1], Carl J. March, Douglas Pat Cerretti, David L. Urdal and Paul J. Conlon

Immunex Corporation, 51 University Street, Seattle, Washington 98101. [1]Present address: Departments of Genetics, SK-50, and Pathology, SM-30, University of Washington, Seattle, WA 98195.

A small hydrophilic peptide of eight amino acids (AspTyrLysAspAspAspAspLys) was engineered onto the N-terminus of a variety of recombinant lymphokines for the purpose of aiding in their detection and purification from yeast supernatants or *E. coli* extracts. An antibody specific for the first four amino acids of this sequence was used as a detection reagent and for immunoaffinity purification of products under mild conditions. Because of the small size of the peptide moiety and its hydrophilic nature, the proteins were unaffected by its presence and retained a high level of biological activity. In addition, it was possible to remove the peptide via an enzymatic cleavage procedure using enterokinase.

Among the many approaches taken to improve the yield and purity of recombinant proteins, one particularly useful procedure is to express the desired polypeptide as part of a larger fusion protein[1,2]. Fusion to a protein-export signal sequence has been used to cause secretion of products from yeast[3] and *E. coli*[4] cells. Furthermore, it has recently been recognized that an attached fusion polypeptide sequence might serve as an aid to identifying or purifying the product. For example, in several cases the added polypeptide segment contains a complete second protein that binds to affinity columns via its specific substrate or ligand. These include β-galactosidase fusion proteins that bind to aminophenylthiogalactosidyl Sepharose columns[5] and protein A fusion proteins that bind to immunoglobulin columns[6]. Such fusion proteins can be highly purified in good yields in a single step by passing cell extracts or supernatants over columns of an appropriate affinity matrix, then eluting the purified fusion protein by changing conditions so that binding is no longer possible. A related approach is to use an antibody directed against the added sequence as a detection or affinity purification reagent[7-9] although the high binding affinity of most antisera and monoclonal antibodies often requires the use of denaturing conditions for elution of the product.

The fusion protein approach as several drawbacks that have not been adequately addressed in the past. First, most fusion protein products fail to fold properly into a native, active state[10]. It is possible that the added polypeptide segment is responsible for this misfolding due to unfavorable interactions during folding of the fusion protein. This often necessitates treatment with strong denaturants such as 8M urea and 7M guanidine HCl followed by refolding procedures[11,12].

A further problem with fusion proteins is that it is often difficult or impossible to remove the additional amino- or carboxyl-terminal sequence from the desired protein product. One solution has been to use relatively drastic conditions and chemical cleavage agents such as CNBr in 70% formic acid[9,12] or low pH incubations[13] to effect cleavage. However, recent studies have attempted to use somewhat milder chemical cleavages such as hydroxylamine treatment at pH 9.0[14] or enzymatic cleavage procedures under physiological conditions. The blood clotting factor X_a has a proteolytic specificity for the tetrapeptide sequence IleGluGlyArg, and has been used to liberate β-globin from a λcII protein fusion sequence[10]. A sequence allowing specific cleavage by collagenase has also been proposed[5]. Sassenfeld and Brewer[15] developed a facilitated ion-exchange purification technique by fusing their proteins to a C-terminal series of arginine residues that are subsequently removed by carboxypeptidase B treatment. These enzymatic processes have been successful in several instances, but often have been limited by poor cleavage yields or by unwanted cleavages that occur within the desired protein sequence[10].

We decided to create a recombinant protein detection and purification system that incorporated several aspects of the above mentioned procedures in order to produce a fusion sequence with a combination of the most desirable properties. Here we report the development of a short, N-terminal fusion sequence AspTyrLysAsp AspAspAspLys, that we refer to as a marker sequence or "Flag™" segment for antibody mediated identification and purification of recombinant proteins. We also describe a monoclonal antibody that reacts with this sequence and can be used as an immuno-affinity purification reagent that purifies marker fusion proteins under very mild conditions. Finally, the marker sequence can be removed by treatment with the protease, enterokinase, which is specific for the five C terminal amino acids of the marker sequence[16]. No harsh treatments are required at any step in this process and proteins purified by this approach retained their biological activity throughout the purification, even while the marker sequence was attached. This paper describes the expression of several such fusion proteins in *Saccharomyces cerevisiae* and *Escherichia coli*.

RESULTS

In order to develop this system, we performed a series of interrelated steps. The eight amino acid marker peptide was engineered onto the N-terminus of the lymphokine interleukin 2 (IL-2)[17] by means of synthetic oligonucleotides. The fusion protein was expressed in yeast and the product purified by conventional means, then used as an immunogen to produce a monoclonal antibody (4E11)

specific for the marker sequence. The antibody functioned as a reagent for a number of different immunochemical procedures including "Westerns," "dot blots," immunoprecipitations, and affinity purification when coupled to a solid support. Furthermore, the discovery that the 4E11 antibody would release its antigen when calcium was removed from the medium led to the development of a mild purification procedure for fusion protein elution from the affinity columns. Next, treatment of the fusion proteins with enterokinase demonstrated that the enzyme was capable of removing the marker segment efficiently, with little or no observable degradation of the desired protein product. Finally, measurements of specific activity demonstrated that, for all proteins tested, no appreciable loss of activity was caused by the presence of the marker segment on the N-termini of the recombinant

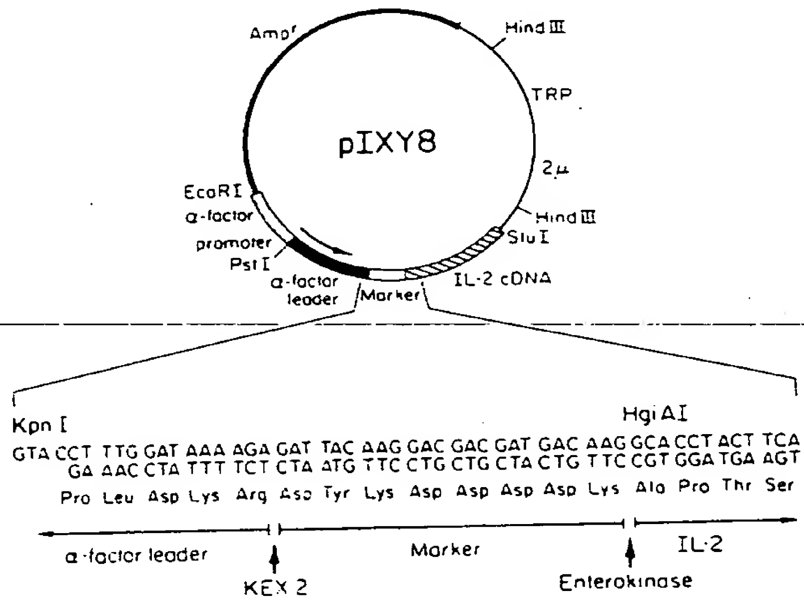

FIGURE 1 Plasmid pIXY8, for expression of the IL-2 fusion protein in yeast. The synthetic oligonucleotides used in constructing this plasmid extended from the KpnI site at the left of the sequence shown below the plasmid diagram, to a blunt end at the HgiAI site near the right side, ending with the codon for the C-terminal lys residue of the marker peptide segment. The IL-2 coding sequence extended from a blunt end before the first codon of IL-2 (Ala) to a StuI site beyond the termination codon. The arrows below the amino acid sequence indicate the sites of cleavage by the KEX2 protease to remove the α-factor precursor sequence from the primary translation product, and by enterokinase, to remove the marker peptide from the product protein.

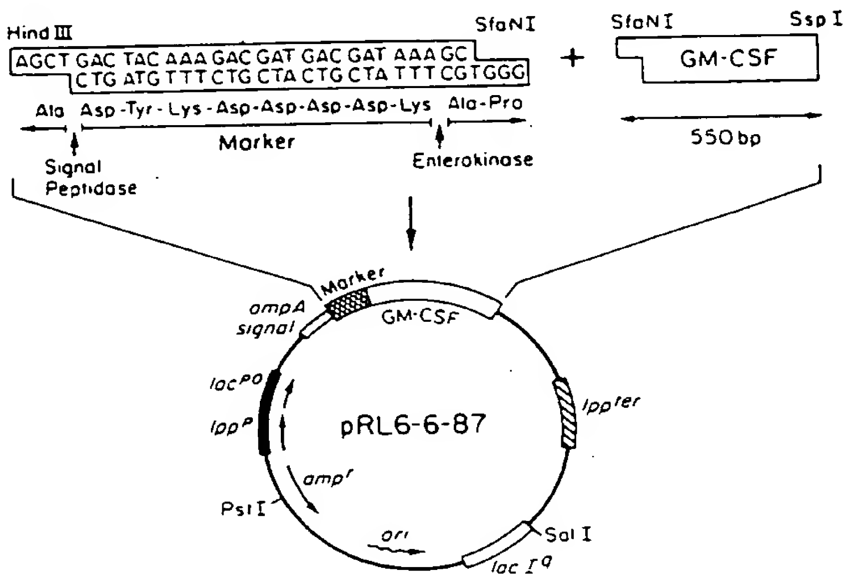

FIGURE 2 Plasmid pRL6-6-87, for expression of the GM-CSF fusion protein in *E. coli*. Abbreviations: *lpp*^{P}, lipoprotein promoter; *lac*^{PO}, lactose promoter-operator region; *ompA* signal, outer membrane protein A signal peptide sequence; GM-CSF, human granulocyte-macrophage colony stimulating factor; *lpp*^{ter}, lipoprotein terminator; *lacI*^{q}, lactose repressor; *ori*, origin of replication.

products.

Secretion of fusion proteins from yeast. Figure 1 shows the plasmid pIXY8, used for production of the IL-2 fusion protein in yeast. Similar constructs with the ADH2 promoter replacing the α-factor promoter were used to express granulocyte colony stimulating factor (G-CSF)[18], interleukin 3 (IL-3)[19], interleukin 4 (IL-4)[20] and granulocyte-macrophage colony stimulating factor (GM-CSF)[21] fusion proteins. All of these proteins were secreted into culture media by yeast. Each had the expected molecular weight for the correctly processed form (i.e. with leader peptide removed) and yielded the expected sequence of the marker peptide on N-terminal amino acid sequence analysis. The IL-2 fusion protein was purified by HPLC for use as an immunogen, while the other fusion proteins were purified by the 4E11 antibody immunoaffinity chromatography procedure (below). For comparison, essentially identical vectors were prepared that expressed each protein without the marker peptide. These products were recovered from the yeast culture media and purified to homogeneity by conventional techniques including HPLC and ion exchange chromatography[22].

Expression of GM-CSF fusion protein in *E. coli*. The construction of the plasmid pRL6-6-87 for expression and secretion of the GM-CSF fusion protein in *E. coli* is outlined in Figure 2. This plasmid allows the secretion of the marker peptide GM-CSF fusion protein by means of the signal peptide from the outer membrane protein OmpA. The product obtained from *E. coli* cultures had the expected molecular weight for the marker peptide GM-CSF fusion and yielded an N-terminal amino acid sequence corresponding to the marker peptide sequence.

The marker-specific antibody. The isotype of the 4E11 antibody is IgG 2B. It was found to be reactive with proteins bearing the marker peptide sequence in a variety of procedures, including ELISAs, dot blots, Western blots, immunoprecipitation and affinity chromatography, as described below. The antibody was found to react with all of the marker peptide fusion proteins that we have produced. The antibody exhibits no reactivity with the non-marker products, or any component present in *E. coli* extracts, or in yeast culture medium.

Purification. Figure 3 shows the results of typical affinity purification chromatograms on affinity columns made with the 4E11 antibody. In Figure 3A, a culture supernatant obtained by fermenting yeast bearing the GM-CSF expression vector was passed over the column to purify the fusion protein that had been secreted into the medium. Medium components were removed by washing with PBS containing 0.5 mM $CaCl_2$, while the GM-CSF remained bound to the antibody. Subsequent elution with PBS containing EDTA dissociated the marker peptide-antibody complex and released the GM-CSF fusion protein as a purified product. The multiple molecular weight species eluting from the column are typical of glycoproteins secreted from yeast and result from heterogeneous glycosylation by yeast cells. All bands were identified as GM-CSF, based on Western analyses using 4E11 as well as anti GM-CSF monoclonal antibody as developing reagents.

Figure 3B shows the results of affinity chromatography of an extract of *E. coli* cells that had been transformed with pRL6-6-87 in order to produce the GM-CSF fusion protein. Chromatography was carried out as with the yeast GM-CSF fusion, except that 1 mM Ca^{+2} was used during washing and 0.1 M glycine HCl pH 3.0 buffer was used to elute the product. The GM-CSF fusion protein eluted as a single molecular weight species because *E. coli* does not glycosylate proteins. The product was essentially pure after this single chromatographic step.

The binding of the 4E11 antibody to the marker peptide is dependent on the presence of calcium. This property has been reported for a few other antibodies in the past[23]. We observed that if insufficient levels of calcium were present in washing buffer, then the fusion proteins would leak from the affinity column, even though they had bound quantitatively when yeast supernatant or *E. coli* extract was passing over the column. By

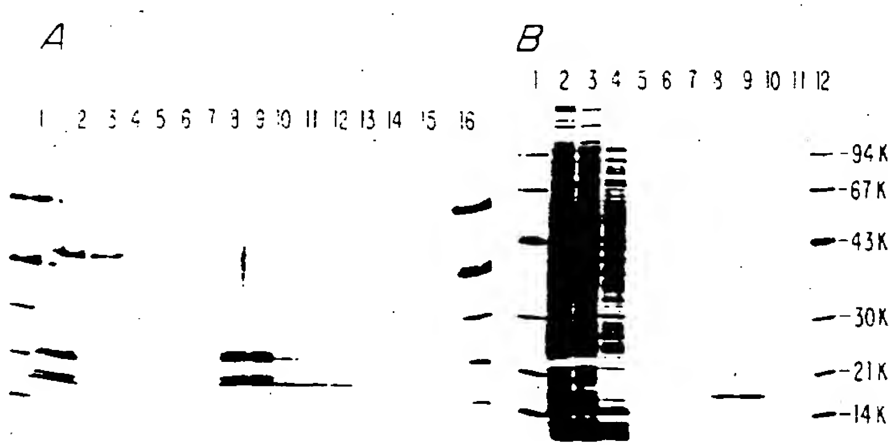

FIGURE 3 Affinity purification of recombinant fusion proteins. Panel A. Silver-stained polyacrylamide gel of yeast GM-CSF fusion protein purification. Lanes are: 1: molecular weight standards; 2: yeast supernatant; 3: flow through material; 4–6: sequential PBS/0.5 mM $CaCl_2$ washes; 7–15: sequential PBS/2.0 mM EDTA elutions; 16: molecular weight standards. Panel B: Silver-stained gel of *E. coli* GM-CSF fusion protein purification. Lanes are: 1: molecular weight standards; 2: *E. coli* supernatant; 3: flow through material; 4–6: sequential PBS/0.5 mM $CaCl_2$ washes; 7–11: sequential 0.1 M Gly-HCl pH 3 elutions; 12: molecular weight standards. Numbers at right indicate M_r values for the standard proteins (in kiloDaltons). The 94K standard was omitted from the gel shown in panel A.

testing the effect of various concentrations of $CaCl_2$ in washing buffer, we determined that concentrations of calcium above 0.3 mM were necessary to retain the fusion proteins on the column. Given this calcium dependence, it was found that rapid elution of fusion proteins could be achieved by using EDTA in the elution buffer. It was also possible to elute proteins simply by using a calcium free elution buffer after the columns had been washed with PBS containing calcium. However, under these conditions the eluted fusion protein tended to spread through more fractions than when the elution buffer contained EDTA.

Enterokinase treatment. Figure 4 shows the results of enterokinase treatment of the IL-2 fusion protein. Increasing amounts of enterokinase were added to identical aliquots of the IL-2 fusion protein, then the samples were incubated for 16 hr at 37°. As the concentration of enzyme was increased, a component appeared at the molecular weight of the authentic IL-2. At the highest concentrations of enzyme, the conversion of fusion protein to authentic protein was complete, and an approximately identical amount of authentic product had been formed. Western blot analyses using 4E11 and anti-IL-2 monoclonal antibodies confirmed the identities of the products seen on the silver gel. Only the higher molecular weight species was reactive with 4E11, but as expected both the higher and lower molecular weight species were reactive with the anti-IL-2 antibody. Amino terminal sequence analysis on the resulting cleavage products indicated that the IL-2 fusion protein was cleaved after the second lysine of the marker peptide yielding the expected sequence for the authentic IL-2 N-terminus. As can be seen in Figure 4, the cleavage is complete, and in this case results in only small amounts of detectable lower molecular weight byproducts formed by enterokinase cleavage within the IL-2 sequence. Essentially the same cleavage pattern was obtained with the other fusions as well. In no case were degradation products present in quantities greater than 10% of total protein (determined by sequence analysis).

Retention of biological activity. The marker peptide fusion proteins expressed in yeast and *E. coli* were all biologically active despite the presence of the marker segment, as is seen in Table 1. In all cases the levels of biological activity obtained with the fusion proteins were comparable to wild-type recombinant proteins expressed without the marker sequence. In the case of GM-CSF the specific activity values in Table 1 were obtained before and after removal of the marker sequence by enterokinase. The nearly identical specific activities obtained indicate that the yield of cleaved product is probably near 100%.

DISCUSSION

The marker peptide fusion system described in this report comprises a unique and widely useful technique for protein identification and purification. In addition, two observations have shown that the marker peptide is compatible with heterologous expression systems: one, both the *E. coli* OmpA signal and the yeast pre-pro alpha factor leader sequences, when fused to the marker peptide, are correctly processed by their respective proteases allowing secretion of marker peptide-protein fusions with the correct N-terminus. And second, because many investigators have reported problems in the N-terminal processing of microbially expressed mammalian proteins[24–26], the ability of the marker peptide to protect the N-terminus of the desired product may be another important feature of this system.

The capability to produce authentic N-termini upon enterokinase treatment is an improvement over a number of other fusion protein approaches. Those that require chemical cleavages using Asp-Pro[12] or Asn-Gly[13] directed reagents, for example, must necessarily leave a proline or glycine at the N-terminus of the product. Although we have not yet tested the ability of enterokinase to cleave the marker sequence from N-termini containing all of the 20 possible amino acids, we have found that it is capable of cleaving products with N-terminal Glu, Ala, Thr, Leu and Ile residues. This suggests that this procedure will be useful for a wide range of N-termini including charged and uncharged, hydrophobic and hydrophilic residues.

There are several requirements that should be met by an efficient detection and purification system based upon fusion polypeptide expression: First, the added marker segment should not interfere with the native folding of proteins to which it is attached. Second, the marker peptide sequence should be intrinsically water soluble and should retain a high degree of exposure in the aqueous environment of the protein, so that it can readily interact with the affinity purification substrate. Third, it should be useful in an affinity purification step that requires only very mild media, and be elutable with a non-denaturing and inexpensive eluant. Finally, the marker peptide should be easy to remove and the product protein should not have any amino acids added or deleted once the marker peptide has been removed. The Flag™ peptide fusion system was designed to possess all of these properties, and our data with several recombinant proteins suggest that it may prove to be a universal purification system for proteins expressed in heterologous organisms.

Several factors were considered in choosing the specific sequence of the marker peptide moiety. We chose to limit the marker peptide sequence to only eight amino acids because it can easily be encoded in a single synthetic oligonucleotide, and because the longest trypsinogen prosequences are of this length. We therefore could be reasonably sure that the trypsin-activating enzyme, enterokinase, would work efficiently to release the peptide. Additionally, because antibodies require up to six or seven amino acids for avid binding interactions, we reasoned that eight amino acids should be the minimum sequence capable of strong binding to an antibody while allowing one or more of the last amino acids on the C-terminal end to act as a spacer to separate the antibody binding portion from interference with the bulk of the protein. Finally, the five C-terminal amino acids of the marker sequence represent the minimal enterokinase specificity site, AspAspAspAspLys.

The choice of Lys at position three of the marker

TABLE 1 Expression level and specific activity of marker fusion proteins.

Protein	mg/l[a]	Specific Activity (U/mg × 10^{-6}) With Marker	Specific Activity (U/mg × 10^{-6}) Without Marker[b]
		Expressed in Yeast	
IL-2	1–3	240 ± 20	280 ± 40
G-CSF	10–20	63 ± 10	12 ± 1.0
IL-3	15–20	49 ± 19	42 ± 12
IL-4	7	5.4 ± 1.6	N.D.[c]
		Expressed in *E. coli*	
GM-CSF[d]	15–20	150 ± 60	140 ± 60

[a]Determined by dot blot assay using 4E11 antibody to detect marker peptide containing material.
[b]Proteins in this column (except *E. coli* GM-CSF) were produced without the marker segment and purified by conventional means.
[c]Not determined.
[d]In this case, specific activities were determined on the same sample before and after enterokinase treatment to remove the marker segment.

sequence causes the marker peptide to contain the hexapeptide sequence, LysAspAspAspAspLys, that has a maximum value on the hydrophilicity scale of Hopp and Woods[27]. Such maximally hydrophilic sequences have been proven to express strong antigenicity and are correspondingly likely to adopt a highly exposed conformation in the three dimensional folding of a protein[28]. As can be seen in Figure 5, it is impossible for any other region of a protein to have a higher hydrophilicity value than this maximally hydrophilic sequence, so the marker segment is virtually guaranteed to be exposed at the surface of any fusion protein. Therefore it can always be expected to be available for binding to antibody. Perhaps most importantly, the strong predilection for externalization should guarantee that the marker segment will not interfere in the adoption of a native conformation by the remainder of the protein.

In addition to the hydrophilic effects of Lys at position three, several other considerations influenced the choice of amino acids at the N-terminus of the marker peptide. Aromatic amino acids have been recognized as major factors in antigen-antibody interactions[29] so a tyrosine was placed at position 2, flanked by charged amino acids. Recent evidence suggests that aromatic residues that are flanked by charged sequences are more likely to be involved in antigenic sites than are other aromatic residues in less polar environments[28]. The decision to place an Asp residue at the N-terminus was made in part because the negative charge on the Asp should aid in exposing the Tyr to antibody, as mentioned above, and in part because, with the inclusion of the Asp at position 1, a total of eight charges are to be found on the marker peptide moiety, including the N-terminal amino group. This preponderance of charged residues was expected to make it likely that antibody binding would be heavily dependent on charge-charge interactions, and therefore might be highly susceptible to elution with commonly used salt solutions such as 2 M $MgCl_2$ or 1 M NaCl[9]. In the end, the serendipitous discovery that Ca^{++} was involved in the charge-dependent binding of the marker sequence to the 4E11 antibody made even these mild salt treatments unnecessary.

We have seen that fusion proteins retain the appropriate specific activity even with the marker segment still attached, and that this activity can be maintained after enterokinase treatment to remove the marker sequence. Comparisons of several of these fusion protein products with their natural counterparts (Table 1) demonstrated that the presence of the marker did not decrease the specific activity of the fusion proteins relative to the same proteins with no extraneous amino acids added. We have recently begun using a larger version of the 4E11 column to prepare proteins in milligram quantities. This level of scale-up required no special procedures or equipment, and can still be done as a bench top experiment. Further scale up for production of gram or kilogram quantities is contemplated, and will be limited only by availability of antibody and enterokinase. One advantage of the 4E11 antibody is that it can be purified on an affinity column comprised of chemically synthesized marker peptide attached to a solid support and eluted with EDTA.

One area that remains potentially problematic is the provision of adequate enterokinase for this process, both in terms of quantity and quality. We sampled commercial sources of enterokinase, but found that our fusion proteins were digested into small fragments, presumably by contaminating chymotrypsin, trypsin and elastase that are likely to be present in these partially purified preparations. Our own crude bovine intestinal preparations also caused substantial unwanted hydrolysis, until we used the protocol of Liepnieks and Light[30] for removing these other proteases. The fact that a minor amount of unwanted cleavage is seen in some cases when the marker peptide is removed (Fig. 4) suggests that traces of contaminating proteolytic activity may still be present in our purified preparations of enterokinase. An ideal solution would be to clone enterokinase and express it in a recombinant organism. This would provide starting material with much lower levels of contaminating proteases, and much higher levels of enterokinase. Furthermore, with the enterokinase gene in hand, it might be possible to engineer a smaller form of the molecule, lacking the hydrophobic portion that binds it to the membranes of the intestinal villi. This would simplify purification and decrease the mass of enterokinase needed for cleavage of the marker segment.

Despite the need for further development of enterokinase, this system for fusion protein detection and purification already represents a useful technique. It offers the possibility of using a single procedure for the purification

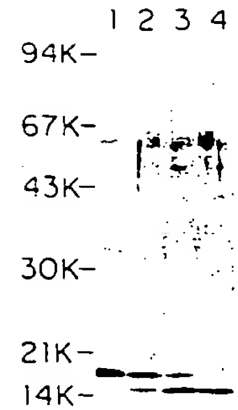

FIGURE 4 Enterokinase digestions. The IL-2 fusion protein was incubated with increasing amounts of bovine enterokinase and the digestion was analyzed by silver-staining PAGE. Lanes are: 1: purified IL-2 fusion protein (200 ng) no enzyme; 2: fusion protein + 2 ng enterokinase; 3: fusion protein + 10 ng enterokinase; 4: fusion protein + 20 ng enterokinase.

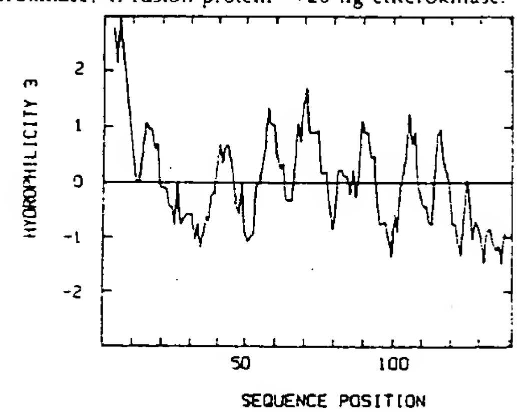

FIGURE 5 Hydrophilicity plot of the IL-2 fusion protein. The profile was generated using the updated HYDRO3 program of Hopp[19]. The scale is oriented so that hydrophilic is at top; hydrophobic is at bottom. Valley regions are expected to be buried portions of the polypeptide, whereas peaks are expected to be exposed at the surface of the protein. The prominent peak near the N-terminus results from the extremely hydrophilic hexapeptide, LysAspAspAspAspLys, contained within the marker peptide sequence.

of multiple fusion proteins. Although it is also possible to purify fusion proteins from yeast medium or *E. coli* cells by conventional means such an ion exchange or reverse phase chromatography, these procedures require new method development for each new protein, whereas with marker fusions, the same process is applicable to all proteins. Finally, because the marker segment does not appear to have decreased the biological activity of any of the proteins that it has been placed on, it may not always be necessary to remove the marker segment in order to obtain an active product. In such cases, this useful "handle" can be retained on the molecule, enabling investigators to readily detect and manipulate their recombinant protein products.

EXPERIMENTAL PROTOCOL

Yeast plasmid construction. The yeast vector used for protein expression has been described previously[11]. This vector contains sequences from pBR322 that allow selection (Amp^r) and replication in *E. coli*, as well as the yeast *Trp1* gene and 2µ origin of replication for selection and autonomous replication in a trp1 yeast strain. Expression of foreign genes is under control of the α-factor promoter[3] or the ADH2 promoter[32] and secretion is directed by the α-factor leader peptide. To generate the IL-2 expression vector pIXY8, the mature coding region of IL-2[17] was fused in-frame to the marker peptide and the α-factor leader by means of a synthetic oligonucleotide linker encoding the five C-terminal amino acids of the α-factor leader and the eight amino acids of the marker peptide (Fig. 1). The vectors that directed the expression of the other products were generated by two modifications of pIXY8. First, the α-factor promoter was removed by restricting the plasmid with EcoRI and PstI, then inserting the ADH2 promoter using a synthetic oligonucleotide linker. Second, the marker and appropriate protein coding sequences were added in place of the IL-2 sequence (Fig. 1) and linked with a synthetic oligonucleotide that extended to the HpaI site.

Growth of yeast strains. *S. cerevisiae* strain XV2181(a/α-trp-1) was grown in either selective medium [YNB $^-$trp, consisting of 0.67% Yeast Nitrogen Base (Difco), 0.5% Casamino acids, 2% glucose, 10 µg/ml adenine and 20 µg/ml uracil] or rich medium (YPD, consisting of 1% Yeast Extract, 2% peptone and 1% glucose supplemented with 80 µg/ml adenine and 80 µg/ml uracil). Yeast transformations were done by selecting for Trp$^+$ transformants[33]. Cultures were grown for biological assay by inoculating 20–50 ml of rich medium with the appropriate strain and growing the cultures at 30°C to stationary phase. Cells were then removed by centrifugation and the medium was filtered through a 0.45µ cellulose acetate filter. Sterile supernates were stored at 4°C. Larger scale fermentations were done in a 10 liter New Brunswick Microferm fermentor. Cells were removed from the medium using a Millipore Pellicon filtration system.

Construction of the *E. coli* vector. Plasmid pIN-III-OmpA$_3$ is an *E. coli* expression vector regulated by the tandem *lpp*^{P} (lipoprotein promoter)/*lac*^{PO} (lactose-promoter-operator) that contains a synthetic *ompA* leader sequence for protein secretion[34]. Construction of the GM-CSF expression vector was accomplished by restriction digestion at the unique BamHI site of pIN-III-OmpA$_3$, followed by its conversion to blunt ends by treatment with reverse transcriptase (Boehringer-Mannheim). The vector was subsequently restricted with HindIII, and used in a three-way ligation with a synthetic oligonucleotide encoding the marker peptide sequence and cDNA encoding GM-CSF to produce pRL6-6-87 as outlined in Figure 2.

Growth of *E. coli*. Plasmid pRL6-6-87 was introduced by transformation into *E. coli* strain JM107, (*Δlac, pro, thi, strA, endA, sbcB15, hspR4*/F', *traD36, proAB*$^-$, *lacI*^{q}-ZΔM15) which was grown at 37°C in M9 minimal medium containing 1% (w/v) methionine assay medium (DIFCO) and ampicillin (50 µg/ml) to an OD_{600} of 0.25. Cultures were induced following addition of isopropyl-β-D-thiogalactopyranoside (IPTG) and cyclic 3'-5' adenosine monophosphate (cAMP) to 2mM and 4mM, respectively, and allowed to accumulate the GM-CSF fusion protein for 2–4 hr. Cells were harvested by centrifugation, and pellets either stored at −70°C or directly processed for extraction and purification of marker fusion proteins. *E. coli* pellets were extracted by the following procedure: the pellet from 500 ml of culture was suspended in 50 ml of 150 mM NaCl, 50 mM NaH_2PO_4, pH8.4, to which 1 mM phenylmethylsulfonylfluoride had been added immediately prior to mixing. After freezing (−70°C) and thawing three times to lyse the *E. coli* cells the sample was incubated at 37°C for 30 min to complete lysis and to extract the protein product. The viscous extract was treated by dounce homogenization to achieve a uniform solution, then centrifuged at 25,000 g for 45 min at 4°C. The supernatant was adjusted to 0.5 mM $CaCl_2$, recentrifuged if necessary to remove any resulting precipitate, then applied to the affinity column.

Preparation of immunogens. Palmitic acid conjugated peptides were produced by solid phase chemical synthesis as described previously[35]. The antigenic marker peptide had the sequence AspTyrLysAspAspAspAspLysGlyProLysLysGly to which palmitoyl moieties had been attached on the epsilon amino groups of the two C-terminal lysines. It is referred to as CDP-marker (C-terminal dipalmitoyl marker peptide). A second palmitoyl peptide, NDP-GM1 (N-terminal dipalmitoyl GM-CSF peptide 1) was used as a non-specific binding control. It had the structure LysGlyGlyGluSerPheLysGluAsnLeuLysAspPheLeuValGly, and also possessed two palmitoyl moieties, in this case attached to the two amino groups of the N-terminal lysine residue. For purification of the IL-2 marker peptide fusion protein, supernatants of yeast expressing the IL-2 fusion protein were applied to a reverse phase HPLC column. The IL-2 fusion product was eluted from the column using a gradient of acetonitrile, as described previously[22].

Immunization. BALB/c female mice were purchased from the Jackson Laboratories (Bar Harbor, ME) and maintained in our animal facility. Mice were immunized subcutaneously with 250 µg of IL-2 fusion protein emulsified in Freund's complete adjuvant, followed with 125 µg of the same protein emulsified in Freund's incomplete adjuvant four weeks later. Two weeks after the second inoculation, a serum antibody titer to IL-2 and the marker peptide was measured by "dot-blot". The animal was then challenged with 10 µg of protein intravenously four days prior to fusion.

Hybridoma derivation. Four days after the intravenous boost, the animals were sacrificed, their spleens removed, and a single cell suspension prepared. The splenocytes were fused to the HAT sensitive myeloma cell, NS-1. The resulting hybridomas were then assayed for the production of antibodies to the marker peptide seven to ten days later by ELISA (see below). One hybridoma antibody consistently produced a positive reaction specifically with the marker peptide moiety. This cell line, designated 4E11, was then cloned by limiting dilution, isotyped, and further characterized.

ELISA. Various peptide solutions (CDP-marker, or NDP-GM1) were applied to HA plates (Millipore, Bedford, MA) at a concentration of 40 ng per well and allowed to incubate for 30 min at room temperature. Nonspecific protein binding sites were blocked by an incubation with 3% bovine serum albumin in Tris buffered saline, pH 7.0 (TBSA) for 1 hr at room temperature. Hybridoma supernatants were added and the plates incubated for 1 hr. Following this incubation, the plates were washed with PBS and an alkaline phosphatase labeled goat anti-mouse antibody (Sigma Chemical, St. Louis, MO) was added. Following a 1 hr incubation, the plates were washed several times with PBS and a colorimetric indicating reagent was added (substrate tablets, Sigma Chemical). Contents of each HA plate were then transferred to a polystyrene 96 well plate (Linbro/Titertek, Flow Laboratories, McLean, VA) and the absorbance at 405 nm determined on a Titerscan (Flow Laboratories).

Production and purification of 4E11 antibody. Pristane-primed BALB/c mice were injected IP with 1 × 10^6 hybridoma cells. Ten to twenty days later, the ascitic fluid was recovered, centrifuged at 1000 × *g* for 30 minutes at 4°C, passed through cotton gauze, and the supernatant stored at −20°C until needed. The monoclonal antibody 4E11 was purified from ascites fluid using MAPS II Protein A Affigel (Bio-Rad, Richmond, CA) affinity chromatography. The purified antibody was found to be homogeneous by SDS-PAGE analysis.

4E11 column preparation. Purified 4E11 immunoglobulin was concentrated by ultrafiltration. After dialysis against 0.1M Hepes buffer, pH 7.5 at 4°C the antibody was coupled to Affigel-10 (Bio-Rad) in accordance with the manufacturer's instructions. A typical antibody-coupled gel contained from 1.5 to 4.5 mg antibody/ml of gel. Columns of 4E11 coupled gel of 1.5 ml bed volume were prepared in polypropylene columns (Bio-Rad) and washed with 15 ml PBS, 15 ml 0.1M glycine HCl, pH 3.0, and stored at 4°C in PBS/0.02% sodium azide.

4E11 column chromatography. Yeast culture filtrates were brought to physiological levels of salt and pH by adding 10X PBS, and made 0.5 mM in $CaCl_2$ by adding 1 M $CaCl_2$, and then loaded onto the 1.5 ml column of 4E11 coupled Affigel 10 under gravity flow. *E. coli* extracts did not require any further additions because

the extraction medium contained physiological levels of salt and pH as well as 0.5 mM $CaCl_2$. Up to 100 ml of filtrate were passed over the column, depending on the level of expression of the recombinant protein. After loading, the column was washed with three to five aliquots of 3 ml of PBS containing 0.5 mM $CaCl_2$. Elution was carried out with PBS lacking $CaCl_2$ and containing 2.0 mM Na_2 EDTA or with 0.1M glycine HCl pH 3.0. Each elution fraction was 1 ml. Yields of purified proteins were determined by amino acid analysis, and were typically 15–40% of the theoretical maximum assuming a 2:1 antigen to antibody binding ratio.

Enterokinase treatment. Enterokinase was purified from bovine intestine by the procedure of Liepnieks and Light[30]. Samples were also provided by A. Light of Purdue University. For enterokinase treatment, fusion proteins eluted from the antibody column were made 10 mM in Tris-HCl (pH 8) and adjusted to pH 8.0 by addition of 1N NaOH. For certain samples, the reaction mixture was made 40 mM in octyl-β-D-glucoside. Following the addition of an appropriate amount of bovine enterokinase (1–10% by weight; typically 0.2–2% by molarity), the reaction mixture was incubated for 16 hours at 37°C. Enterokinase dilutions were made from a 1 mg/ml stock solution of enzyme in 10 mM Tris-HCl, pH 8 kept at −70°C.

Bioassays. The activity of IL-2 was measured using the murine IL-2 dependent T-cell line CTLL-2[36]. The activity of GM-CSF was measured in a human bone marrow proliferation assay[37] and the activity of IL-3 was measured by FDC-P2 cell proliferation[38]. IL-4 and G-CSF were assayed as described[39,40]. Specific activities were derived by measuring the biological activities of purified samples of each protein, after quantifying by amino acid analysis.

Acknowledgments

The authors wish to thank David Amberg, Janet Merriam, Aline Akelis, Norman Boiani, Bruce Hess, and David McMasters for excellent technical assistance. We would also like to thank Linda Troup for preparation of the manuscript.

Received 19 April 1988; accepted 5 July 1988.

References

1. Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F., and Boyer, H. W. 1977. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. Science **198**:1056–1063.
2. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., and Riggs, A. D. 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. Proc. Natl. Acad. Sci. USA **76**:106–110.
3. Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenback, G. T., Urdea, M. S., Valenzuela, P., and Barr, P. J. 1984. α-Factor-directed synthesis and secretion of foreign proteins in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **81**:4642–4646.
4. Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L., and Gilbert, W. 1978. A bacterial clone synthesizing proinsulin. Proc. Natl. Acad. Sci. USA **75**:3727–3731.
5. Germino, J. and Bastia, D. 1984. Rapid purification of a cloned gene product by genetic fusion and site-specific proteolysis. Proc. Natl. Acad. Sci. USA **81**:4692–4696.
6. Abrahmsen, L., Moks, T., Nilsson, B., and Uhlen, M. 1986. Secretion of heterologous gene products to the culture medium of *Escherichia coli*. Nucleic Acids Res. **14**:7487–7500.
7. Munro, S. and Pelham, H. R. B. 1984. Use of peptide tagging to detect proteins expressed from cloned genes. EMBO J. **3**:3087–3093.
8. Ikehara, M., Ohtsuka, E., Tokunaga, T., Nishikawa, S., Uesugi, S., Tanaka, T., Aoyama, Y., Kikyodani, S., Fujimoto, K., Yanase, K., Fuchimura, K., and Morioka, H. 1986. Inquiries into the structure-function relationship of ribonuclease T1 using chemically synthesized coding sequences. Biochemistry **85**:4695–4699.
9. Wilchek, M., Miron, T., and Kohn, J. 1984. Affinity chromatography. Methods in Enzymol. **104**:3–55.
10. Nagai, K. and Thogerson, H. C. 1987. Synthesis and sequence specific proteolysis of hybrid proteins produced in *Escherichia coli*. Methods Enzymol. **153**:461–481.
11. Schoner, R. G., Ellis, L. F., and Schoner, B. E. 1985. Isolation and purification of protein granules from *Escherichia coli* cells overproducing bovine growth hormone. Bio/Technology **3**:151–154.
12. Sung, W. L., Yao, F. L., Zahab, D. M., and Narang, S. A. 1986. Short synthetic oligodeoxyribonucleotide leader sequences enhance accumulation of human proinsulin synthesized in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83**:561–565.
13. Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L., and Uhlen, M. 1985. Efficient secretion and purification of human insulin-like growth factor I with a gene fusion vector in *Staphylococci*. Nucleic Acids Res. **13**:1151–1162.
14. Moks, T., Abrahmsen, L., Osterlof, B., Josephson, S., Ostling, M., Enfors, S., Persson, I., Nilsson, B., and Uhlen, M. 1987. Large-scale affinity purification of human insulin-like growth factor I from culture medium of *Escherichia coli*. Bio/Technology **5**:379–382.
15. Sassenfeld, H. M. and Brewer, S. J. 1984. A polypeptide fusion designed for the purification of recombinant proteins. Bio/Technology **2**:76–81.
16. Maroux, S., Baratti, J., and Desnuelle, P. 1971. Purification and specificity of procine enterokinase. J. Biol. Chem. **246**:5031–5039.
17. Taniguchi, T., Matsui, H., Fujita, T., Takaoka, C., Kashima, N., Yoshimoto, R., and Hamuro, J. 1983. Structure and expression of a cloned cDNA for human interleukin-2. Nature **302**:305–310.
18. Nagata, S., Tsuchiya, M., Asano, S., Kaziro, Y., Yamazaki, T., Yamamoto, O., Hirata, Y., Kubota, N., Oheda, M., Nomura, H., and Ono, M. 1986. Molecular cloning and expression of cDNA for human granulocyte colony stimulating factor. Nature **319**:415–417.
19. Yang, Y-C., Ciarletta, A. B., Temple, P. A., Chung, M. P., Kovacic, S., Witek-Giannotti, J. S., Leary, A. C., Kriz, R., Donahue, R. E., Wong, G. G., and Clark, S. C. 1986. Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. Cell **47**:3–10.
20. Yokota, T., Otsuka, T., Mosmann, T., Banchereau, J., DeFrance, T., Blanchard, D., DeVries, J. E., Lee, F., and Arai, K-I. 1986. Isolation and characterization of a human cDNA clone, homologous to mouse B-cell stimulatory factor 1, that expresses B-cell and T-cell stimulating activities. Proc. Natl. Acad. Sci. USA **83**:5894–5898.
21. Cantrell, M. A., Anderson, D., Cerretti, D. P., Price, V., McKereghan, K., Tushinski, R. J., Mochizuki, D. Y., Larsen, A., Grabstein, K., Gillis, S., and Cosman, D. 1985. Cloning, sequence, and expression of a human granulocyte/macrophage colony stimulating factor. Proc. Natl. Acad. Sci. USA **82**:6250–6254.
22. Urdal, D. L., Mochizuki, D., Conlon, P. J., March, C. J., Remerowski, M. L., Eisenman, J., Ramthun, C., and Gillis, S. 1984. Lymphokine purification by reversed-phase high performance liquid chromatography. J. Chromatography **296**:171–179.
23. Maurer, P. H., Clark, L. G., and Liberti, P. A. 1970. Antigenicity of polypeptides (polyaamino acids): Calcium-dependent and independent antibodies. J. Immunol. **105**:567–573.
24. Hitzeman, R. A., Leung, D. W., Perry, L. J., Kohr, W. J., Levine, H. L., and Goeddel, D. V. 1983. Secretion of human interferons by yeast. Science **219**:620–625.
25. Aston, R., Cooper, L., Holder, A., Ivanyi, J., and Preece, M. 1985. Monoclonal antibodies to human growth hormone can distinguish between pituitary and genetically engineered forms. Molec. Immunol. **22**:271–275.
26. Kronheim, S. R., Cantrell, M. A., Deeley, M. C., March, C. J., Glackin, P. J., Anderson, D. M., Hemenway, T., Merriam, J. E., Cosman, D., and Hopp, T. P. 1986. Purification and characterization of human interleukin-1 expressed in *Escherichia coli*. Bio/Technology **4**:1078–1082.
27. Hopp, T. P. and Woods, K. R. 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
28. Hopp, T. P. 1986. Protein surface analysis: Methods for identifying antigenic determinants and other interaction sites. J. Immunol. Methods **88**:1–18.
29. Sela, M. and Arnon, R. 1960. Studies on the chemical basis of the antigenicity of proteins. Biochem. J. **75**:91–102.
30. Liepnieks, J. J. and Light, A. 1979. The preparation and properties of bovine enterokinase. J. Biol. Chem. **254**:1677–1683.
31. Price, V., Mochizuki, D., March, C. J., Cosman, D., Deeley, M. C., Klinke, R., Clevenger, W., Gillis, S., Baker, P., and Urdal, D. 1987. Expression, purification and characterization of recombinant murine granulocyte-macrophage colony-stimulating factor and bovine interleukin-2 from yeast. Gene **55**:287–293.
32. Beier, D. R. and Young, E. T. 1982. Characterization of a regulatory region upstream of the ADR2 locus of *S. cerevisiae*. Nature **300**:724–728.
33. Hinnen, A., Hicks, J. B., and Fink, G. R. 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
34. Libby, R. T., Braedt, G., Kronheim, S. R., March, C. J., Urdal, D. L., Chiaverotti, T. A., Tushinski, R. J., Mochizuki, D. Y., Hopp, T. P., and Cosman, D. 1987. Expression and purification of native human granulocyte-macrophage colony stimulating factor from an *Escherichia coli* vector. DNA **6**:221–229.
35. Hopp, T. P. 1984. Immunogenicity of a synthetic HBsAg peptide. Enhancement by conjugation to a fatty acid carrier. Molec. Immunol. **21**:13–16.
36. Gillis, S., Ferm, M. M., Ou, W., and Smith, K. 1978. T cell growth factor: parameters of production and a quantitative microassay for activity. J. Immunol. **120**:2027–2033.
37. Mochizuki, D. Y., Eisenman, J., Conlon, P. J., Park, L. S., and Urdal, D. L. 1986. Development and characterization of antiserum to murine granulocyte-macrophage colony stimulating factor. J. Immunol. **136**:3760–3709.
38. Prestidge, R. L., Watson, J. D., Urdal, D. L., Mochizuki, D., Conlon, P., and Gillis, S. 1984. Biochemical comparison of murine colony stimulating factors secreted by a T cell lymphoma and a myelomonocytic leukemia. J. Immunol. **133**:293–298.
39. Grabstein, K., Eisenman, J., Mochizuki, D., Shanebeck, K., Conlon, P., Hopp, T., March, C., and Gillis, S. 1986. Purification to homogeneity of B-cell stimulating factor: A molecule that stimulates proliferation of multiple lymphokine-dependent cell lines. J. Exp. Med. **163**:1405–1414.
40. Metcalf, D. 1986. The molecular biology and functions of granulocyte-macrophage colony stimulating factors. Blood **67**:257–267.